C01462740

Secret NOFORN

DECLASSIFIED UNDER AUTHORITY OF THE
INTERAGENCY SECURITY CLASSIFICATION APPEALS PANEL,
E.O. 13526, SECTION 5.3(b)(3)

ISCAP APPEAL NO. 2014-004, document no. 1
DECLASSIFICATION DATE: September 19, 2016

THE CENTRAL INTELLIGENCE AGENCY AND OVERHEAD RECONNAISSANCE

The U-2 and OXCART Programs, 1954 - 1974

Gregory W. Pedlow and
Donald E. Welzenbach

Secret

SECRET

Warning Notice

Intelligence Sources or Methods Involved (WNINTEL)

National Security Information

Unauthorized Disclosure Subject to Criminal Sanctions

Dissemination Control Abbreviations

NOFORN (NF)	Not releasable to foreign nationals
NOCONTRACT (NC)	Not releasable to contractors or contractor/consultants
PROPIN (PR)	Caution—proprietary information involved
ORCON (OC)	Dissemination and extraction of information controlled by originator
REL...	This information has been authorized for release to...
WN	WNINTEL—Intelligence sources and methods involved
	Classified by [redacted] Declassify: OADR Derived from multiple sources

Withheld under statutory authority of the Central Intelligence Agency Act of 1949 (50 U.S.C., section 403g)

All material on this page is Unclassified.

C01462740

The Central Intelligence Agency and Overhead Reconnaissance:

The U-2 and OXCART Programs, 1954-1974

C01462740

~~SECRET~~

~~Secret~~ NOFORN

The Central Intelligence Agency and Overhead Reconnaissance:

The U-2 and OXCART Programs, 1954-1974

Gregory W. Pedlow
and
Donald E. Welzenbach

History Staff
Central Intelligence Agency
Washington, D.C.

1992

~~Secret~~

~~SECRET~~

Contents

Chapter 1
Searching for a System

The Need for High-Altitude Reconnaissance 1
Early Postwar Aerial Reconnaissance 2
New Approaches to Photoreconnaissance 4
The Air Force Search for a New Reconnaissance Aircraft 8
Lockheed CL-282 Supporters and the CIA 13
Scientists and Overhead Reconnaissance 17
The BEACON HILL Report 17
Concern About the Danger of a Soviet Surprise Attack 19
The Air Force Intelligence Systems Panel 21
British Overflight of Kapustin Yar 23
The Intelligence Systems Panel and the CL-282 24
The Technological Capabilities Panel 26
Project Three Support for the Lockheed CL-282 27
A Meeting With the President 32
CIA and Air Force Agreement on the CL-282 33

Chapter 2
Developing the U-2

The Establishment of the U-2 Project 39
Funding Arrangements for Project AQUATONE 43
Major Design Features of the U-2 45
The Development of the Camera System 48
Preparations for Testing the U-2 56
Security for the U-2 Project 59
The CIA – Air Force Partnership 60
Technical Challenges to High-Altitude Flight 61
Delivery of the First U-2 66
Initial Testing of the U-2 68
U-2s, UFOs, and Operation BLUE BOOK 72

Hiring U-2 Pilots 73
Pilot Training 75
Final Tests of the U-2 76
Three Fatal Crashes in 1956 79
Coordination of Collection Requirements 80
Preparations To Handle the Product of U-2 Missions 82
The Impact of the Air Force Project GENETRIX Balloons 84
AQUATONE Briefings for Selected Members of Congress 88
The U-2 Cover Story 89

Chapter 3
U-2 Operations in the Soviet Bloc and Middle East, 1956-1958

The Deployment of Detachment A to Lakenheath 94
The Move to Wiesbaden 95
President Eisenhower's Attitude Toward Overflights 96
First Overflights of Eastern Europe 100
First U-2 Flights Over the Soviet Union 104
Soviet Protest Note 109
The End of the Bomber Gap 111
Tactical Intelligence From U-2s During the Suez Crisis 112
Renewed Overflights of the Soviet Union 122
Radar-Deceptive "Dirty Birds" 128
The New Detachment C 133
Detachment B Flights From Pakistan 135
The Decline of Detachment A 139
Cooperation With Norway 142
Declining Overflight Activity 143
Concerns About Soviet Countermeasures Against the U-2 147
More Powerful Engines for the U-2 149
Intervention in Lebanon, 1958 152
British Participation in the U-2 Project 153
The U-2 Project at the Beginning of 1959 157

Chapter 4
The Final Overflights of the Soviet Union, 1959-1960

The U-2 and the "Missile-Gap" Debate 159
The Last Overflight: Operation GRAND SLAM 170
The Aftermath of the U-2 Downing 177
The Withdrawal of the Overseas Detachments 181
The Fate of Francis Gary Powers 183
Changes in Overflight Procedures After May 1960 187

Chapter 5
U-2 Operations After May 1960

U-2 Operations in Latin America 197
U-2 Support to the Bay of Pigs Invasion 197
Aerial Refueling Capability for the U-2 198
U-2 Coverage During the Cuban Missile Crisis 199
U-2s Over South America 211
U-2 Operations in Asia 211
Detachment C and the Indonesian Revolt of 1958 211
China Offshore Islands Dispute of 1958 215
U-2 Support for DDP Operations in Tibet 216
U-2Cs for Detachment C 217
U-2 Crash in Thailand 219
End of Detachment C Operations 219
Detachment G Missions Over Laos and North Vietnam 221
New Detachment on Taiwan 222
Use of Detachment H Aircraft by US Pilots 230
U-2s in India 231
Increasing Responsibilities, Inadequate Resources in Asia 233
Advanced ECM Equipment for Detachment H 237
Use of Infrared Scanner Over PRC Nuclear Plants 238
[redacted] 240
The End of U-2 Overflights of Mainland China 242
Peripheral Missions by Detachment H 244
Operation SCOPE SHIELD Over North Vietnam 246
Improvements in U-2 Technology 247
Modification of U-2s for Aircraft Carrier Deployment 247
Use of Carrier-Based U-2 To Film a French Nuclear Test Site 249
A New Version of the U-2 251
Replacement of the Original U-2s With U-2Rs 253
The Final Years of the U-2 253
Support to Other Agencies 254
Overseas Deployment Exercises and Missions 255
The Phaseout of the Office of Special Activities 257

25X1, E.O.13526

Chapter 6
The U-2's Intended Successor: Project OXCART, 1956-1968

The Evaluation of Designs for a Successor to the U-2 260
Competition Between Lockheed and Convair 267
The Selection of the Lockheed Design 270
Efforts To Reduce the A-12's Radar Cross Section 274
The OXCART Contract 277

New Technologies Necessitated By OXCART's High Speed 279
Designing the OXCART's Cameras 281
Choosing Pilots for OXCART 283
Selection of a Testing Site for the OXCART 283
Delivery of the First OXCART 286
Changes in the Project Management 286
OXCART's First Flights 288
Speed-Related Problems 290
New Versions of the OXCART 291
The Question of Surfacing a Version of the OXCART 292
Additional Problems During Final Testing 295
Discussions on the OXCART's Future Employment 297
First A-12 Deployment: Operation BLACK SHIELD 304
The End of the OXCART Program 307
Possible Successors to the OXCART 312
Summary of the OXCART Program 313

Chapter 7
Conclusion

U-2 Overflights of the Soviet Union 315
Participation of Allies in the U-2 Program 319
U-2s as Collectors of Tactical Intelligence 319
Advances in Technology 320
Cooperation With the Air Force 321
Impact of the Overhead Reconnaissance Program on the CIA 321

Appendix A: Acronyms 325

Appendix B: Key Personnel 327

Appendix C: Electronic Devices Carried by the U-2 335

Appendix D: U-2 Overflights of the Soviet Union, 4 July 1954–1 May 1960 337

Appendix E: Unmanned Reconnaissance Projects 339

Bibliography 347

Index 355

C01462740

FOREWORD

This History Staff Monograph offers a comprehensive and authoritative history of the CIA's manned overhead reconnaissance program, which from 1954 to 1974 developed and operated two extraordinary aircraft, the U-2 and the A-12 OXCART. It describes not only the program's technological and bureaucratic aspects, but also its political and international context. The manned reconnaissance program, along with other overhead systems that emerged from it, changed the CIA's work and structure in ways that were both revolutionary and permanent. The formation of the Directorate of Science and Technology in the 1960s, principally to develop and direct reconnaissance programs, is the most obvious legacy of the events recounted in this study.

The authors tell an engrossing story. The struggle between the CIA and the US Air Force to control the U-2 and A-12 OXCART projects reveals how the manned reconnaissance program confronted problems that still beset successor programs today. The U-2 was an enormous technological success: its first flight over the USSR in July 1956 made it immediately the most important source of intelligence on the Soviet Union. Using it against the Soviet target it was designed for nevertheless produced a persistent tension between its program managers and the President. The program managers, eager for coverage, repeatedly urged the President to authorize frequent missions over the Soviet Union. President Eisenhower, from the outset doubtful of the prudence and propriety of invading Soviet airspace, only reluctantly allowed any overflights at all. After the Soviets shot down Francis Gary Powers' U-2 on 1 May 1960, President Eisenhower forbade any further U-2 flights over the USSR. Since the Agency must always assess a covert operation's potential payoff against the diplomatic or military cost if it fails, this account of the U-2's employment over the Soviet Union offers insights that go beyond overhead reconnaissance programs.

Indeed, this study should be useful for a variety of purposes. It is the only history of this program based upon both full access to CIA records and extensive classified interviews of its participants. The authors have found records that were nearly irretrievably lost and have interviewed participants whose personal recollections gave information available nowhere else. Although the story of the manned

reconnaissance program offers no tidy model for imitation, it does reveal how resourceful managers coped with unprecedented technological challenges and their implications for intelligence and national policy. For this reason, the program's history provides profitable reading for intelligence professionals and policymakers today.

Many people made important contributions to the production of this volume. In the History Staff's preparation of the manuscript, Gerald Haines did the final revision, [redacted] again demonstrated her high talent as a copy editor, and [redacted] provided staunch secretarial support throughout. As usual, we are indebted to more members than we can name from the Publications, Design, and Cartography Centers in the Office of Current Production and Analytic Support, whose lively interest in the publication went far beyond the call of duty. Their exceptional professional skill and the masterly work of the Printing and Photography Group combined to create this handsome volume.

Withheld under statutory authority of the Central Intelligence Agency Act of 1949 (50 U.S.C., section 403g)

Donald E. Welzenbach, who began this study, and Gregory W. Pedlow, who completed it, brought complementary strengths to this work. A veteran of CIA service since 1960, Mr. Welzenbach began research on this study in 1983, when he joined the DCI History Staff on a rotational assignment from the Directorate of Science and Technology. After tireless documentary research and extensive interviewing, he finished a draft manuscript of the history before returning to his directorate. In early 1986, Gregory W. Pedlow, a new member of the DCI History Staff, was assigned to complete the study. A Johns Hopkins University Ph.D. who has served as an Army intelligence officer and University of Nebraska professor of history, Dr. Pedlow undertook important research in several new areas, and reorganized, edited, and revised the entire manuscript before leaving CIA to become NATO Historian in late 1989. The final work, which has greatly benefited from both authors' contributions, is the CIA's own history of the world's first great overhead reconnaissance program. ◆

J. Kenneth McDonald
Chief, CIA History Staff

April 1992

C01462740

PREFACE

When the Central Intelligence Agency came into existence in 1947, no one foresaw that, in less than a decade, it would undertake a major program of overhead reconnaissance, whose principal purpose would be to fly over the Soviet Union. Traditionally, the military services had been responsible for overhead reconnaissance, and flights deep into unfriendly territory only took place during wartime. By the early 1950s, however, the United States had an urgent and growing need for strategic intelligence on the Soviet Union and its satellite states. At great risk, US Air Force and Navy aircraft had been conducting peripheral reconnaissance and shallow-penetration overflights, but these missions were paying a high price in lives lost and increased international tension. Furthermore, many important areas of the Soviet Union lay beyond the range of existing reconnaissance aircraft. The Air Force had therefore begun to develop a high-altitude reconnaissance aircraft that would be able to conduct deep-penetration reconnaissance missions over the Soviet Union. President Dwight D. Eisenhower and his civilian scientific advisers feared that the loss of such an aircraft deep in Soviet territory could lead to war and therefore authorized the development of new nonmilitary aircraft, first the U-2 and later the A-12 OXCART, to be manned by civilians and operated only under cover and in the greatest secrecy. Primary responsibility for this new reconnaissance program was assigned to the Central Intelligence Agency, but the Air Force provided vital support.

The Agency's manned overhead reconnaissance program lasted 20 years. It began with President Eisenhower's authorization of the U-2 project in late 1954 and ended with the transfer of the remaining Agency U-2s to the Air Force in 1974. During this period the CIA developed a successor to the U-2, the A-12 OXCART, but this advanced aircraft saw little operational use and the program was canceled in 1968 after the Air Force deployed a fleet of similar aircraft, a military variant of the A-12 called the SR-71.

Neither of these aircraft remains secret today. A great deal of information about the U-2 and its overflight program became known to the public after 1 May 1960, when the Soviet Union shot down a CIA U-2 and publicly tried its pilot, Francis Gary Powers. Four years

later, at press conferences in February and July 1964, President Lyndon B. Johnson revealed the existence of the OXCART-type of aircraft, although only in its military YF-12A (interceptor) and SR-71 (strategic reconnaissance) versions.

The two CIA reconnaissance aircraft have also been the subject of a number of books, beginning with David Wise's and Thomas B. Ross's *The U-2 Affair* in 1962 and then Francis Gary Powers' memoirs, *Operation Overflight*, in 1970. Two recent books give many more details about the U-2 and OXCART aircraft: Michael Beschloss's *Mayday: Eisenhower, Khrushchev and the U-2 Affair* (1986) and William Burrows's *Deep Black: Space Espionage and National Security* (1987). Although well written and generally accurate, these books suffer from their authors' lack of access to classified official documentation. By drawing upon the considerable amount of formerly classified data on the U-2 now available to the public, Beschloss has provided an accurate and insightful depiction of the U-2 program in the context of the Eisenhower administration's overall foreign policy, but his book does contain errors and omissions on some aspects of the U-2 program. Burrows's broader work suffers more from the lack of classified documentation, particularly in the OXCART/SR-71 section, which concentrates on the Air Force aircraft because little information about the Agency's aircraft has been officially declassified and released.

After the present study of the Agency's overhead reconnaissance projects was completed, a new book on the U-2 was published in the United Kingdom. Chris Pocock's *Dragon Lady: The History of the U-2 Spyplane* is by far the most accurate unclassified account of the U-2 program. Pocock has been able to compensate for his lack of access to classified documents by interviewing many former participants in the program, especially former pilots. Pocock is also quite familiar with aircraft itself, for he had worked with Jay Miller on the latter's excellent technical study of the U-2: *Lockheed U-2* (1983).

There has also been a classified official study of the U-2 and OXCART programs. In 1969 the Directorate of Science and Technology published a *History of the Office of Special Activities* by

Helen Hill Kleyla and Robert D. O'Hern. This 16-volume Top Secret Codeword study of the Agency's reconnaissance aircraft provides a wealth of technical and operational information on the two projects but does not attempt to place them in their historical context. Without examining the international situation and bureaucratic pressures affecting the president and other key policymakers, however, it is impossible to understand the decisions that began, carried out, and ended the CIA's reconnaissance aircraft projects.

In preparing this study of CIA's overhead reconnaissance program, the authors drew on published sources, classified government documents, and interviews with key participants from the CIA, Air Force, contractors, scientific advisory committees, and the Eisenhower administration. The interviews were particularly important for piecing together the story of how the CIA became involved in overhead reconnaissance in the first place because Agency documentation on the prehistory of the U-2 project is very sketchy and there are no accurate published accounts. Research on the period of actual reconnaissance operations included the records of the Director of Central Intelligence, the Office of Special Activities in the Directorate of Science and Technology, and the Intelligence Community Staff, along with documents from the Eisenhower Presidential Library in Abilene, Kansas, and additional interviews.

Both authors are grateful for the assistance they have received from many individuals who played important roles in the events they recount. Without their help a good deal of this story could never have become known. The assistance of Agency records management officers in the search for documents on the overhead reconnaissance program is also greatly appreciated.

To ensure that this study of the Agency's involvement in overhead reconnaissance reaches the widest possible audience, the authors have kept it at the Secret classification level. As a result, some aspects of the overhead reconnaissance program, particularly those involving satellites and related interagency agreements, have had to be described in very general terms. The omission of such information is not significant for this book, which focuses on the Agency's reconnaissance aircraft. ◆

Searching for a System

THE NEED FOR HIGH-ALTITUDE RECONNAISSANCE

For centuries, soldiers in wartime have sought the highest ground or structure in order to get a better view of the enemy. At first it was tall trees, then church steeples and bell towers. By the time of the American Civil War and the Franco-Prussian War of 1870-71, observers were using hot-air balloons to get up in the sky for a better view of the "other side of the hill." With the advent of dry film, it became possible to carry cameras into the sky to record the disposition of enemy troops and emplacements. Indeed, photoreconnaissance proved so valuable during World War I that in 1938 Gen. Werner von Fritsch, Commander in Chief of the German Army, predicted: "The nation with the best aerial reconnaissance facilities will win the next war." [1]

By World War II, lenses, films, and cameras had undergone many improvements, as had the airplane, which could fly higher and faster than the primitive craft of World War I. Now it was possible to use photoreconnaissance to obtain information about potential targets before a bombing raid and to assess the effectiveness of the bombing afterward.

Peacetime applications of high-altitude photography at first included only photomapping and surveying for transcontinental highways and mineral and oil exploration. There was little thought given to using photography for peacetime espionage until after World War II, when the Iron Curtain rang down and cut off most forms of communication between the Soviet Bloc of nations and the rest of the world.

[1] Roy M. Stanley II, *World War II Photo Intelligence* (New York: Scribners, 1981), p. 16.

By 1949 the Soviet Union and the states of Eastern Europe had been effectively curtained off from the outside world, and the Soviet military carried out its planning, production, and deployment activities with the utmost secrecy. All Soviet strategic capabilities—bomber forces, ballistic missiles, submarine forces, and nuclear weapons plants—were concealed from outside observation. The Soviet air defense system, a prime consideration in determining US retaliatory policies, was also largely an unknown factor.

Tight security along the Soviet Bloc borders severely curtailed the movement of human intelligence sources. In addition, the Soviet Union made its conventional means of communication—telephone, telegraph, and radio-telephone—more secure, thereby greatly reducing the intelligence available from these sources. The stringent security measures imposed by the Communist Bloc nations effectively blunted traditional methods for gathering intelligence: secret agents using covert means to communicate intelligence, travelers to and from target areas who could be asked to keep their eyes open and report their observations later, wiretaps and other eavesdropping methods, and postal intercepts. Indeed, the entire panoply of intelligence tradecraft seemed ineffective against the Soviet Bloc, and no other methods were available.

Early Postwar Aerial Reconnaissance

Although at the end of World War II the United States had captured large quantities of German photos and documents on the Soviet Union, this material was rapidly becoming outdated. The main source of current intelligence on the Soviet Union's military installations was interrogation of prisoners of war returning from Soviet captivity. To obtain information about Soviet scientific progress, the intelligence community established several programs to debrief German scientists who had been taken to the Soviet Union after the end of the war but were now being allowed to leave.[2]

[2] At the end of World War II, the British had established Project DRAGON to gain information from German scientists who had worked on the Peenemunde rocket project, and the term DRAGON later was used to refer to individuals possessing scientific or technical information. In 1948 the US Air Force set up Project WRINGER in Germany to gather intelligence on the Soviet Union from defectors and refugees; this project was later absorbed into the combined armed forces/CIA Defector Reception Center (DRC), which began operations in February 1951. In October 1951, a separate organization to exploit individuals with scientific or technical backgrounds, especially German scientists who had worked inside the Soviet Union, came into existence. This organization was known as the Returnee Exploitation Group (REG) and was located in Frankfurt. By 1958 the flow of scientists was so small that the REG merged with the DRC. [redacted] *The Defector Reception Center Germany, 1951 to 1967*, Clandestine Service Historical Series CSHP-41 (CIA: History Staff, 1972), pp. 5-6, 29-30 ~~(S)~~.

Withheld under statutory authority of the Central Intelligence Agency Act of 1949 (50 U.S.C., section 403g)

Interrogation of returning Germans offered only fragmentary information, and this source could not be expected to last much longer. As a result, in the late 1940s, the US Air Force and Navy began trying to obtain aerial photography of the Soviet Union. The main Air Force effort involved Boeing RB-47 aircraft (the reconnaissance version of the B-47 jet-propelled medium bomber) equipped with cameras and electronic "ferret" equipment that enabled aircrews to detect tracking by Soviet radars. At that time the Soviet Union had not yet completely ringed its borders with radars; and much of the interior also lacked radar coverage. Thus, when the RB-47s found a gap in the air-warning network, they would dart inland to take photographs of any accessible targets. These "penetration photography" flights (called SENSINT—sensitive intelligence—missions) occurred along the northern and Pacific coasts of Russia. One RB-47 aircraft even managed to fly 450 miles inland and photograph the city of Igarka in Siberia. Such intrusions brought protests from Moscow but no Soviet military response.[3]

In 1950 there was a major change in Soviet policy. Air defense units became very aggressive in defending their airspace, attacking all aircraft that came near the borders of the Soviet Union. On 8 April 1950, Soviet fighters shot down a US Navy Privateer patrol aircraft over the Baltic Sea. Following the outbreak of the Korean war in June 1950, the Soviet Union extended its "severe air defense policy" to the Far East. In the autumn of 1951, Soviet aircraft downed a twin-engine US Navy Neptune bomber near Vladivostok. An RB-29 lost in the Sea of Japan on 13 June 1952 was probably also a victim of Soviet fighters. The United States was not the only country affected by the new aggressive Soviet air defense policy; Britain and Turkey also reported attacks on their planes.[4]

---

[3] A. L. George, *Case Studies of Actual and Alleged Overflights*, 1930-1953, Rand Study RM-1349 (Santa Monica: Rand, 1955) (S). Arthur S. Lundahl and Dino Brugioni, interview by Donald E. Welzenbach, tape recording, Washington, DC, 14 December 1983 (~~TS~~ Codeword). Recordings, transcripts, and notes for the interviews conducted for this study are on file at the DCI History Staff.

[4] Jeffrey Richelson states on page 121 of *American Espionage and the Soviet Target* (New York: Morrow, 1987) that "the first recorded attack by Soviet air defense forces, in this case fighters, occurred on October 22, 1949." In this incident, however, Soviet fighters did not attempt to hit the US aircraft; they merely fired warning shots. The real change in Soviet policy did not occur until the April 1950 downing of the US Navy Privateer. George, *Case Studies*, pp. 1-2, 6, 9-16 (S).

The Soviet Union's air defense policy became even more aggressive in August 1952, when its reconnaissance aircraft began violating Japanese airspace over Hokkaido, the northernmost Japanese home island. Two months later, on 7 October 1952, Soviet fighter aircraft stalked and shot down a US RB-29 flying over Hokkaido. Aerial reconnaissance of the Soviet Union and surrounding areas had become a very dangerous business.

Despite the growing risks associated with aerial reconnaissance of the Soviet Bloc, senior US officials strongly believed that such missions were necessary. The lack of information about the Soviet Union, coupled with the perception that it was an aggressive nation determined to expand its borders—a perception that had been greatly strengthened by the Soviet-backed North Korean invasion of South Korea in June 1950—increased US determination to obtain information about Soviet intentions and capabilities and thus reduce the danger of being surprised by a Soviet attack.

New Approaches to Photoreconnaissance

While existing Navy and Air Force aircraft were flying their risky reconnaissance missions over the Soviet Union, the United States began planning for a more systematic and less dangerous approach using new technology. One of the leading advocates of the need for new, high-altitude reconnaissance aircraft was Richard S. Leghorn, a Massachusetts Institute of Technology graduate and employee of Eastman Kodak who had commanded the Army Air Forces' 67th Reconnaissance Group in Europe during World War II. After the war he returned to Kodak but maintained his interest in photoreconnaissance. Leghorn strongly believed in the need for what he called pre-D-day reconnaissance, that is, reconnaissance of a potential enemy before the outbreak of actual hostilities, in contrast to combat reconnaissance in wartime. In papers presented in 1946 and 1948, Leghorn argued that the United States needed to develop such a capability, which would require high-altitude aircraft and high-resolution cameras. The outbreak of the Korean war gave Leghorn an opportunity to put his ideas into effect. Recalled to active duty by the Air Force, Lieutenant Colonel Leghorn became the head of the Reconnaissance Systems Branch of the Wright Air Development Command at Dayton, Ohio, in April 1951.[5]

[5] Richard S. Leghorn, interview by Donald E. Welzenbach, tape recording, Washington, DC, 19 August 1985 ~~(S)~~.

In Leghorn's view, altitude was the key to success for overhead reconnaissance. Since the best Soviet interceptor at that time, the MIG-17, had to struggle to reach 45,000 feet,[6] Leghorn reasoned that an aircraft that could exceed 60,000 feet would be safe from Soviet fighters. Recognizing that the fastest way to produce a high-altitude reconnaissance aircraft was to modify an existing aircraft, he began looking for the highest flying aircraft available in the Free World. This search soon led him to a British twin-engine medium bomber—the Canberra—built by the English Electric Company. The Canberra had made its first flight in May 1949. Its speed of 469 knots (870 kilometers per hour) and its service ceiling of 48,000 feet made the Canberra a natural choice for high-altitude reconnaissance work. The Royal Air Force quickly developed a reconnaissance version of the Canberra, the PR3 (the PR stood for photoreconnaissance), which began flying in March 1950.[7]

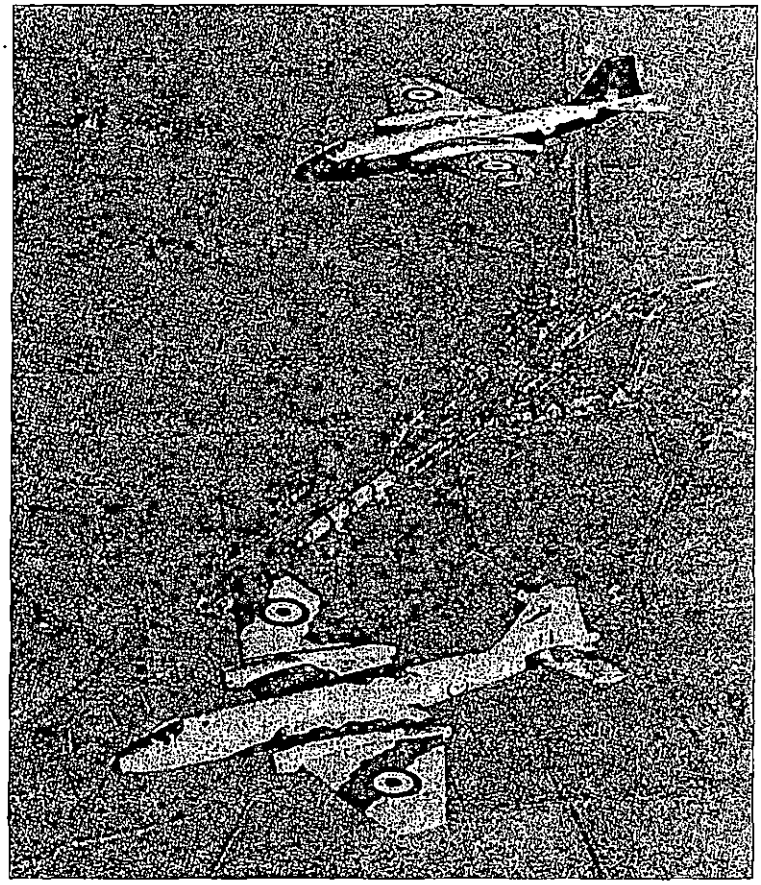
Richard S. Leghorn

At Leghorn's insistence, the Wright Air Development Command invited English Electric representatives to Dayton in the summer of 1951 to help find ways to make the Canberra fly even higher. By this time the Air Force had already adopted the bomber version of the Canberra, which the Glenn L. Martin Aircraft Company was to produce under license as the B-57 medium bomber. Leghorn and his English Electric colleagues designed a new Canberra configuration with very long high-lift wings, new Rolls-Royce Avon-109 engines, a solitary pilot, and an airframe that was stressed to less than the standard military specifications. Leghorn calculated that a Canberra so equipped might reach 63,000 feet early in a long mission and as high as 67,000 feet as the declining fuel supply lightened the aircraft. He believed that such a modified Canberra could penetrate the Soviet Union and China for a radius of 800 miles from bases around their periphery and photograph up to 85 percent of the intelligence targets in those countries.

Leghorn persuaded his superiors to submit his suggestion to the Pentagon for funding. He had not, however, cleared his idea with the Air Research and Development Command, whose reconnaissance

[6] 13,716 meters. To avoid giving a false impression of extremely precise measurements, original English measuring system figures in round numbers have not been converted to the metric system. To convert feet to meters, multiply by 0.3048. To convert airspeeds in knots (nautical miles per hour) to kilometers per hour, multiply by 1.85.

[7] Dick van der Aart, *Aerial Espionage, Secret Intelligence Flights by East and West* (Shrewsbury, England: Airlife Publishing, 1985), p. 18.



RAF Canberra Mark-PR3

division in Baltimore, headed by Lt. Col. Joseph J. Pellegrini, had to approve all new reconnaissance aircraft designs. Pellegrini's unit reviewed Leghorn's design and ordered extensive modifications. According to Leghorn, Pellegrini was not interested in a special-purpose aircraft that was only suitable for covert peacetime reconnaissance missions, for he believed that all Air Force reconnaissance aircraft should be capable of operating under wartime conditions. Pellegrini therefore insisted that Leghorn's design meet the specifications for combat aircraft, which required heavily stressed airframes, armor plate, and other apparatus that made an aircraft too heavy to reach the higher altitudes necessary for safe overflights of the Soviet Bloc. The final result of Leghorn's concept after its alteration by Pellegrini's staff was the RB-57D in 1955, whose maximum altitude

was only 64,000 feet. Meanwhile Leghorn, frustrated by the rejection of his original concept, had transferred to the Pentagon in early 1952 to work for Col. Bernard A. Schriever, Assistant for Development Planning to the Air Force's Deputy Chief of Staff for Development.[8]

In his new position Leghorn became responsible for planning the Air Force's reconnaissance needs for the next decade. He worked closely with Charles F. (Bud) Wienberg—a colleague who had followed him from Wright Field—and Eugene P. Kiefer, a Notre Dame–educated aeronautical engineer who had designed reconnaissance aircraft at the Wright Air Development Center during World War II. All three of these reconnaissance experts believed that the Air Force should emphasize high-altitude photoreconnaissance.

Underlying their advocacy of high-altitude photoreconnaissance was the belief that Soviet radars would not be able to track aircraft flying above 65,000 feet. This assumption was based on the fact that the Soviet Union used American-built radar sets that had been supplied under Lend-Lease during World War II. Although the SCR-584 (Signal Corps Radio) target-tracking radar could track targets up to 90,000 feet, its high power consumption burned out a key component quickly, so this radar was normally not turned on until an early warning radar had detected a target. The SCR-270 early warning radar could be left on for much longer periods and had a greater horizontal range (approximately 120 miles) but was limited by the curvature of the earth to a maximum altitude of 40,000 feet. As a result, Leghorn, Kiefer, and Wienberg believed that an aircraft that could ascend to 65,000 feet before entering an area being swept by the early warning radar would go undetected, because the target-tracking radars would not be activated.[9]

The problem with this assumption was that the Soviet Union, unlike Britain and the United States, had continued to improve radar technology after the end of World War II. Even after evidence of improved Soviet radar capabilities became available, however, many advocates of high-altitude overflight continued to believe that aircraft flying above 65,000 feet were safe from detection by Soviet radars.

---

[8] Leghorn interview (S).

[9] Ivan A. Getting, interview by Donald E. Welzenbach, Los Angeles, 28 August 1988 (U).

The Air Force Search for a New Reconnaissance Aircraft

With interest in high-altitude reconnaissance growing, several Air Force agencies began to develop an aircraft to conduct such missions. In September 1952, the Air Research and Development Command gave the Martin Aircraft Company a contract to examine the high-altitude potential of the B-57 by modifying a single aircraft to give it long, high-lift wings and the American version of the new Rolls-Royce Avon-109 engine. These were the modifications that Richard Leghorn had suggested during the previous year.[10]

At about the same time, another Air Force office, the Wright Air Development Command (WADC) in Dayton, Ohio, was also examining ways to achieve sustained flight at high altitudes. Working with two German aeronautical experts—Woldemar Voigt and Richard Vogt—who had come to the United States after World War II, Air Force Maj. John Seaberg advocated the development of a new aircraft that would combine the high-altitude performance of the latest turbojet engines with high-efficiency wings in order to reach ultrahigh altitudes. Seaberg, an aeronautical engineer for the Chance Vought Corporation until his recall to active duty during the Korean war, was serving as assistant chief of the New Developments Office of WADC's Bombardment Branch.

By March 1953, Seaberg had expanded his ideas for a high-altitude aircraft into a complete request for proposal for "an aircraft weapon system having an operational radius of 1,500 nm [nautical miles] and capable of conducting pre- and post-strike reconnaissance missions during daylight, good visibility conditions." The requirement stated that such an aircraft must have an optimum subsonic cruise speed at altitudes of 70,000 feet or higher over the target, carry a payload of 100 to 700 pounds of reconnaissance equipment, and have a crew of one.[11]

The Wright Air Development Command decided not to seek proposals from major airframe manufacturers on the grounds that a smaller company would give the new project a higher priority and

[10] Philip G. Strong, Chief, Operations Staff, OSI, Memorandum for the Record, "Reconnaissance Capabilities," 21 August 1953, OSI records (S).

[11] Jay Miller, *Lockheed U-2*, Aerograph 3 (Austin, Texas: Aerofax, 1983), p. 10.

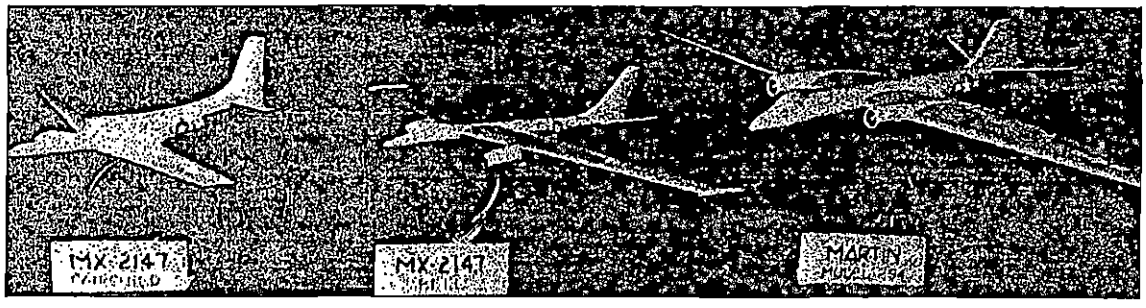
produce a better aircraft more quickly. In July 1953, the Bell Aircraft Corporation of Buffalo, New York, and the Fairchild Engine and Airplane Corporation of Hagerstown, Maryland, received study contracts to develop an entirely new high-altitude reconnaissance aircraft. In addition, the Glenn L. Martin Company of Baltimore was asked to examine the possibility of improving the already exceptional high-altitude performance of the B-57 Canberra. By January 1954 all three firms had submitted their proposals. Fairchild's entry was a single-engine plane known as M-195, which had a maximum altitude potential of 67,200 feet; Bell's was a twin-engine craft called the Model 67 (later the X-16), which had a maximum altitude of 69,500 feet; and Martin's design was a big-wing version of the B-57 called the Model 294, which was to cruise at 64,000 feet. In March 1954, Seaberg and other engineers at Wright Field, having evaluated the three contending designs, recommended the adoption of both the Martin and Bell proposals. They considered Martin's version of the B-57 an interim project that could be completed and deployed rapidly while the more advanced concept from Bell was still being developed.[12]

Air Force headquarters soon approved Martin's proposal to modify the B-57 and was very much interested in the Bell design. But word of the competition for a new reconnaissance airplane had reached another aircraft manufacturer, the Lockheed Aircraft Corporation, which submitted an unsolicited design.

Lockheed had first become aware of the reconnaissance aircraft competition in the fall of 1953. John H. (Jack) Carter, who had recently retired from the Air Force to become the assistant director of Lockheed's Advanced Development Program, was in the Pentagon on business and dropped in to see Eugene P. Kiefer, an old friend and colleague from the Air Force's Office of Development Planning (more commonly known as AFDAP from its Air Force office symbol). Kiefer told Carter about the competition for a high-flying aircraft and expressed the opinion that the Air Force was going about the search in the wrong way by requiring the new aircraft to be suitable for both strategic and tactical reconnaissance.

Immediately after returning to California, Carter proposed to Lockheed Vice President L. Eugene Root (previously the top civilian official in the Air Force's Office of Development Planning) that

[12] The request for proposal, known as "Design Study Requirements, Identification No. 53WC-16507," has been reprinted in Miller, *Lockheed U-2*, pp. 10-11.



Designs for the Air Force competition for a high-altitude reconnaissance aircraft

Lockheed also submit a design. Carter noted that the proposed aircraft would have to reach altitudes of between 65, 000 and 70,000 feet and correctly forecast, "If extreme altitude performance can be realized in a practical aircraft at speeds in the vicinity of Mach 0.8, it should be capable of avoiding virtually all Russian defenses until about 1960." Carter added, "To achieve these characteristics in an aircraft which will have a reasonably useful operational life during the period before 1960 will, of course, require very strenuous efforts and extraordinary procedures, as well as nonstandard design philosophy." Some of the "nonstandard" design characteristics suggested by Carter were the elimination of landing gear, the disregard of military specifications, and the use of very low load factors. Carter's memorandum closed with a warning that time was of the essence: "In order that this special aircraft can have a reasonably long and useful life, it is obvious that its development must be greatly accelerated beyond that considered normal." [13]

Lockheed's senior officials approved Carter's proposal, and early in 1954 the corporation's best aircraft designer—Clarence L. (Kelly) Johnson—began working on the project, then known as the CL-282 but later to become famous under its Air Force designator—the U-2. Already one of the world's leading aeronautical engineers, Kelly Johnson had many successful military and civilian designs to his credit, including the P-38, P-80, F-104, and Constellation. Johnson quickly came up with a radical design based upon the fuselage of the F-104 jet fighter but incorporating a high-aspect-ratio sailplane wing. To save weight and thereby increase the aircraft's altitude, Johnson decided to stress the airframe to only 2.5 units of

[13] Miller, *Lockheed U-2*, p. 12.

gravity (g's) instead of the military specification strength of 5.33 g's. For the power plant he selected the General Electric J73/GE-3 nonafterburning turbojet engine with 9,300 pounds of thrust (this was the same engine he had chosen for the F-104, which had been the basis for the U-2 design).[14] Many of the CL-282's design features were adapted from gliders. Thus, the wings and tail were detachable. Instead of a conventional landing gear, Johnson proposed using two skis and a reinforced belly rib for landing—a common sailplane technique—and a jettisonable wheeled dolly for takeoff. Other features included an unpressurized cockpit and a 15-cubic-foot payload area that could accommodate 600 pounds of sensors. The CL-282's maximum altitude would be just over 70,000 feet with a 2, 000-mile range. Essentially, Kelly Johnson had designed a jet-propelled glider.[15]

Kelly Johnson

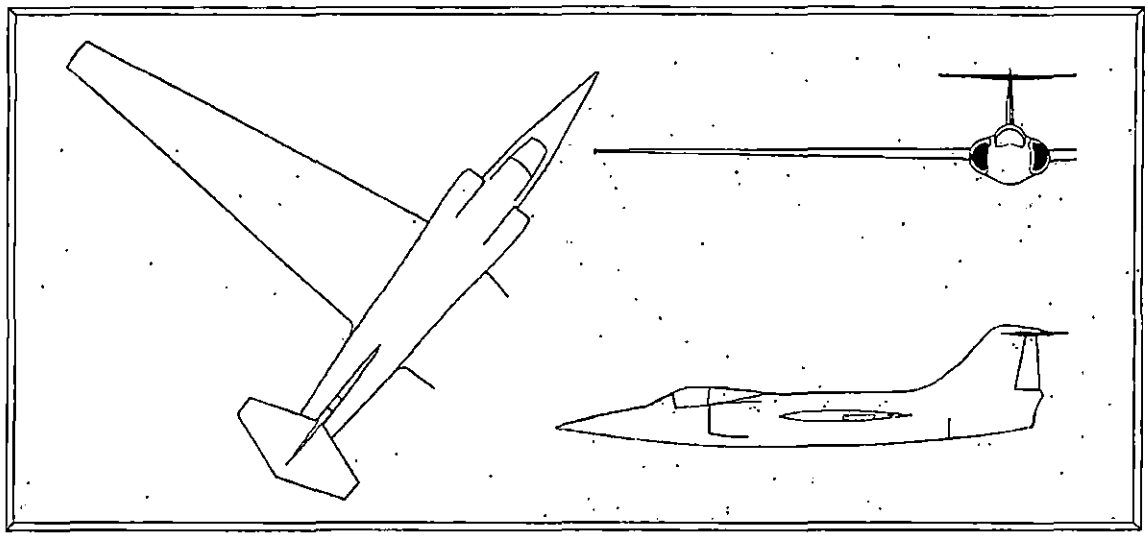
Early in March 1954, Kelly Johnson submitted the CL-282 design to Brig. Gen. Bernard A. Schriever's Office of Development Planning. Eugene Kiefer and Bud Wienberg studied the design and recommended it to General Schriever, who then asked Lockheed to submit a specific proposal. In early April, Kelly Johnson presented a full description of the CL-282 and a proposal for the construction and maintenance of 30 aircraft to a group of senior Pentagon officials that included Schriever's superior, Lt. Gen. Donald L. Putt, Deputy Chief of Staff for Development, and Trevor N. Gardner, Special Assistant for Research and Development to the Secretary of the Air Force. Afterward Kelly Johnson noted that the civilian officials were very much interested in his design but the generals were not.[16]

The CL-282 design was also presented to the commander of the Strategic Air Command (SAC), Gen. Curtis E. LeMay, in early April by Eugene Kiefer, Bud Wienberg, and Burton Klein from the Office of

[14] Lockheed Corporation, "Strategic Reconnaissance and Intelligence," Development Planning Note #1, 30 November 1953 (U).

[15] Miller, *Lockheed U-2*, p. 12. For more details on Kelly Johnson's original proposal, see "Profile of CL-282 High Altitude Aircraft prepared by Lockheed Aircraft Corporation, 5 March 1954" in Helen H. Kleyla and Robert D. O'Hern, *History of the Office of Special Activities*, DS&T, Directorate of Science and Technology Historical Series OSA-1, 16 vols. (CIA: DS&T, 1969), chap. 1, annex 2 (~~TS~~ Codeword). The 16 volumes of this history contain 20 chapters, each paginated separately. Future references will be shortened to *OSA History*, followed by the relevant chapter and page numbers.

[16] Kelly Johnson Papers, "Log for Project X," April 1954, Lockheed Corporation, Advanced Development Projects Division, Burbank, California.



The Lockheed CL-282

Development Planning. According to Wienberg, General LeMay stood up halfway through the briefing, took his cigar out of his mouth, and told the briefers that, if he wanted high-altitude photographs, he would put cameras in his B-36 bombers and added that he was not interested in a plane that had no wheels or guns. The general then left the room, remarking that the whole business was a waste of his time.[17]

Meanwhile, the CL-282 design proceeded through the Air Force development channels and reached Major Seaberg at the Wright Air Development Command in mid-May. Seaberg and his colleagues carefully evaluated the Lockheed submission and finally rejected it in early June. One of their main reasons for doing so was Kelly Johnson's choice of the unproven General Electric J73 engine. The engineers at Wright Field considered the Pratt and Whitney J57 to be the most powerful engine available, and the designs from Fairchild, Martin, and Bell all incorporated this engine. The absence of conventional landing gear was also a perceived shortcoming of the Lockheed design.[18]

Another factor in the rejection of Kelly Johnson's submission was the Air Force preference for multiengine aircraft. Air Force reconnaissance experts had gained their practical experience during

[17] C. F. Wienberg, telephone conversation with Donald E. Welzenbach, 23 July 1988 (U).

[18] Miller, *Lockheed U-2*, p. 12.

World War II in multiengine bombers. In addition, aerial photography experts in the late 1940s and early 1950s emphasized focal length as the primary factor in reconnaissance photography and, therefore, preferred large aircraft capable of accommodating long focal-length cameras. This preference reached an extreme in the early 1950s with the development of the cumbersome 240-inch Boston camera, a device so large that the YC-97 Boeing Stratocruiser that carried it had to be partially disassembled before the camera could be installed. Finally, there was the feeling shared by many Air Force officers that two engines are always better than one because, if one fails, there is a spare to get the aircraft back to base. In reality, however, aviation records show that single-engine aircraft have always been more reliable than multiengine planes. Furthermore, a high-altitude reconnaissance aircraft deep in enemy territory would have little chance of returning if one of the engines failed, forcing the aircraft to descend.[19]

On 7 June 1954, Kelly Johnson received a letter from the Air Force rejecting the CL-282 proposal because it had only one engine and was too unusual and because the Air Force was already committed to the modification of the Martin B-57.[20] By this time, the Air Force had also selected the Bell X-16; the formal contract calling for 28 aircraft was signed in September. Despite the Air Force's selection of the X-16, Lockheed continued to work on the CL-282 and began seeking new sources of support for the aircraft.

Lockheed CL-282 Supporters and the CIA

Although the Air Force's uniformed hierarchy had decided in favor of the Bell and Martin aircraft, some high-level civilian officials continued to favor the Lockheed design. The most prominent proponent of the Lockheed proposal was Trevor Gardner, Special Assistant for Research and Development to Air Force Secretary Harold E. Talbott. Gardner had many contacts in west coast aeronautical circles because before coming to Washington he had headed the Hycon Manufacturing Company, which made aerial cameras in Pasadena, California. He had been present at Kelly Johnson's presentation on the CL-282 at the Pentagon in early April 1954 and believed that this

[19] Allen F. Donovan, interview by Donald E. Welzenbach, Corona del Mar, California, 20 May 1985 (S).

[20] Johnson, "Log for Project X," 7 June 1954.

Trevor Gardner

design showed the most promise for reconnaissance of the Soviet Union. This belief was shared by Gardner's special assistant, Frederick Ayer, Jr., and Garrison Norton, an adviser to Secretary Talbott.[21]

According to Norton, Gardner tried to interest SAC commander LeMay in the Lockheed aircraft because Gardner envisioned it primarily as a collector of strategic, rather than tactical, intelligence. But General LeMay had already shown that he was not interested in an unarmed aircraft. Gardner, Ayer, and Norton then decided to seek CIA support for the high-flying aircraft. At that time the Agency's official involvement in overhead reconnaissance was limited to advising the Air Force on the problems of launching large camera-carrying balloons for reconnaissance flights over hostile territory (for the details of this program, see chapter 2). The Chief of the Operations Staff in the Office of Scientific Intelligence, Philip G. Strong, however, served on several Air Force advisory boards and kept himself well informed on developments in reconnaissance aircraft.[22]

Gardner, Norton, and Ayer met with Strong in the Pentagon on 12 May 1954, six days before the Wright Air Development Command began to evaluate the Lockheed proposal. Gardner described Kelly Johnson's proposal and showed the drawings to Strong. After this meeting, Strong summarized his impressions of the Air Force's search for a high-altitude reconnaissance aircraft:

> *Proposals for special reconnaissance aircraft have been received in the Air Staff from Lockheed, Fairchild, and Bell. . . . The Lockheed proposal is considered to be the best. It has been given the type designation of CL-282 and in many respects is a jet-powered glider based essentially on the Lockheed Day Fighter XF-104. It is primarily subsonic but can attain transonic speeds over the target with a consequent loss of range. With an altitude of 73,000 feet over the target it has a combat radius of 1,400 nautical miles. . . . The CL-282 can be manufactured*

[21] Garrison Norton, interview by Donald E. Welzenbach, tape recording, Washington, DC, 23 May 1983 (S); Michael R. Beschloss, *Mayday: Eisenhower, Khrushchev and the U-2 Affair* (New York: Harper & Row, 1986), p. 79.

[22] Strong was a colonel in the Marine Corps Reserve and often used that title even though he was not on active duty. He later advanced to the rank of brigadier general in the reserve. For Strong's contacts with senior Air Force officials concerning the CL-282, see the Norton interview (S).

mainly with XF-104 jigs and designs. . . . The prototype of this plane can be produced within a year from the date of order. Five planes could be delivered for operations within two years.

The Bell proposal is a more conventional aircraft having normal landing gear. As a result, its maximum altitude over target is 69,500 feet and the speed and range are not as good as the Lockheed CL-282.[23]

Philip Strong

Gardner's enthusiasm for the CL-282 had given Strong the false impression that most Air Force officials supported the Lockheed design. In reality, the Air Force's uniformed hierarchy was in the process of choosing the modified version of the Martin B-57 and the new Bell X-16 to meet future reconnaissance needs.

During their meeting with Strong, Trevor Gardner, Frederick Ayer, and Garrison Norton explained that they favored the CL-282 because it gave promise of flying higher than the other designs and because at maximum altitude its smaller radar cross section might make it invisible to existing Soviet radars. The three officials asked Strong if the CIA would be interested in such an aircraft. Strong promised to talk to the Director of Central Intelligence's newly hired Special Assistant for Planning and Coordination, Richard M. Bissell, Jr., about possible Agency interest in the CL-282.[24]

Richard Bissell had already had an active and varied career before he joined the CIA. A graduate of Groton and Yale, Bissell studied at the London School of Economics for a year and then completed a doctorate at Yale in 1939. He taught economics, first at Yale and then from 1942 at the Massachusetts Institute of Technology (MIT), where he became a full professor in 1948. During World War II, Bissell had managed American shipping as executive officer of the Combined Shipping Adjustment Board. After the war, he served as deputy director of the Marshall Plan from 1948 until the end of 1951, when he became a staff member of the Ford Foundation. His first association with the Agency came in late 1953, when he undertook a contract study of possible responses the United

[23] Philip G. Strong, Memorandum for the Record, "Special Aircraft for Penetration Photo Reconnaissance," 12 May 1954, OSI records (now in OSWR), job 80R-01424, box 1 (S).

[24] Karl H. Weber, *The Office of Scientific Intelligence, 1949-68*, Directorate of Science and Technology Historical Series OSI-1 (CIA: DS&T, 1972), vol. I, tab A, pp. 16-17 (TS Codeword).

Richard M. Bissell, Jr.

States might use against the Soviet Bloc in the event of another uprising such as the East Berlin riots of June 1953. Bissell quickly concluded that there was not much hope for clandestine operations against Bloc nations. As he remarked later: "I know I emerged from that exercise feeling that very little could be done." This belief would later make Bissell a leading advocate of technical rather than human means of intelligence collection.[25]

Bissell joined the Agency in late January 1954 and soon became involved in coordination for the operation aimed at overthrowing Guatemalan President Jacobo Arbenz. He was, therefore very preoccupied when Philip Strong approached him in mid-May 1954 with the concept of the proposed spyplane from Lockheed. Bissell said that the idea had merit and told Strong to get some topflight scientists to advise on the matter. Afterward he returned to the final planning for the Guatemalan operation and promptly forgot about the CL-282.[26]

Meanwhile, Strong went about drumming up support for high-altitude overflight. In May 1954 he persuaded DCI Allen W. Dulles to ask the Air Force to take the initiative in gaining approval for an overflight of the Soviet guided-missile test range at Kapustin Yar. Dulles's memorandum did not mention the CL-282 or any of the other proposed high-altitude aircraft. CIA and Air Force officials met on several occasions to explore the overflight proposal, which the Air Force finally turned down in October 1954.[27]

Although Allen Dulles was willing to support an Air Force overflight of the Soviet Union, he was not enthusiastic about the CIA undertaking such a project. Few details about Dulles's precise attitude toward the proposed Lockheed reconnaissance aircraft are available, but many who knew him believe that he did not want the CIA to become involved in projects that belonged to the military, and the Lockheed CL-282 had been designed for an Air Force requirement.

---

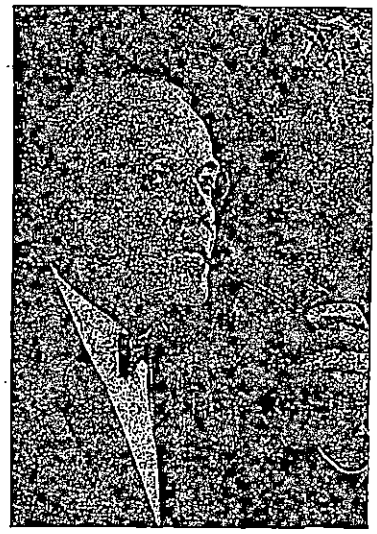
[25] Thomas Powers, *The Man Who Kept the Secrets: Richard Helms and the CIA* (New York: Alfred A. Knopf, 1979), p. 79; Beschloss, *Mayday*, pp. 86-89.

[26] Memorandum for H. Marshall Chadwell, Assistant Director/Scientific Intelligence, from Chief, Support Staff, OSI, "Review of OSA Activities Concerned with Scientific and Technical Collection Techniques," 13 May 1955, p. 6, OSI (OSWR) records, job 80R-01424, box 1 (S); Richard M. Bissell, Jr., interview by Donald E. Welzenbach, tape recording, Farmington, Connecticut, 8 November 1984 (~~S~~).

[27] Memorandum for Richard M. Bissell, Special Assistant to the Director for Planning and Coordination, from Philip G. Strong, Chief, Operations Staff, OSI, "Overflight of Kapustin Yar," 15 October 1954, OSI (OSWR) records, job 80R-01424, box 1 (~~TS~~, downgraded to ~~S~~).

Moreover, high-altitude reconnaissance of the Soviet Union did not fit well into Allen Dulles's perception of the proper role of an intelligence agency. He tended to favor the classical form of espionage, which relied on agents rather than technology.[28]

At this point, the summer of 1954, Lockheed's CL-282 proposal still lacked official support. Although the design had strong backers among some Air Force civilians and CIA officials, the key decisionmakers at both Air Force and CIA remained unconvinced. To make Kelly Johnson's revolutionary design a reality, one additional source of support was necessary: prominent scientists serving on government advisory boards.



DCI Allen W. Dulles

SCIENTISTS AND OVERHEAD RECONNAISSANCE

Scientists and engineers from universities and private industry had played a major role in advising the government on technical matters during World War II. At the end of the war, most of the scientific advisory boards were disbanded, but within a few years the growing tensions of the Cold War again led government agencies to seek scientific advice and assistance. In 1947 the Air Force established a Scientific Advisory Board, which met periodically to discuss topics of current interest and advise the Air Force on the potential usefulness of new technologies. The following year the Office of Defense Mobilization established the Scientific Advisory Committee, but the Truman administration made little use of this new advisory body.[29]

The BEACON HILL Report

In 1951 the Air Force sought even more assistance from scientists because the Strategic Air Command's requests for information about targets behind the Iron Curtain could not be filled. To look for new ways of conducting reconnaissance against the Soviet Bloc, the Air Force's Deputy Chief of Staff for Development, Maj. Gen. Gordon P. Saville, added 15 reconnaissance experts to an existing project on air

[28] Powers, *Man Who Kept the Secrets*, pp. 103-104; Edwin H. Land, interview by Donald E. Welzenbach, tape recording, Cambridge, Massachusetts, 17 and 20 September 1984 (TS Codeword); Robert Amory, Jr., interview by Donald E. Welzenbach and Gregory W. Pedlow, Washington, DC, 22 April 1987 (~~S~~).

[29] For more information on the Air Force's use of scientists see Thomas A. Sturm, *The USAF Scientific Advisory Board: Its First Twenty Years, 1944-1964* (Washington, DC: USAF Historical Office, 1967) (U).

defense known as Project LINCOLN, then under way at the Massachusetts Institute of Technology. By the end of the year, these experts had assembled in Boston to begin their research. Their headquarters was located over a secretarial school on Beacon Hill, which soon became the codename for the reconnaissance project. The consultants were called the BEACON HILL Study Group.

The study group's chairman was Kodak physicist Carl F. P. Overhage, and its members included James G. Baker and Edward M. Purcell from Harvard; Saville Davis from the *Christian Science Monitor*; Allen F. Donovan from the Cornell Aeronautical Laboratory; Peter C. Goldmark from Columbia Broadcasting System Laboratories; Edwin H. Land, founder of the Polaroid Corporation; Stewart E. Miller of Bell Laboratories; Richard S. Perkin of the Perkin-Elmer Company; and Louis N. Ridenour of Ridenour Associates, Inc. The Wright Air Development Command sent Lt. Col. Richard Leghorn to serve as its liaison officer.[30]

During January and February 1952, the BEACON HILL Study Group traveled every weekend to various airbases, laboratories, and firms for briefings on the latest technology and projects. The panel members were particularly interested in new approaches to aerial reconnaissance, such as photography from high-flying aircraft and camera-carrying balloons. One of the more unusual (albeit unsuccessful) proposals examined by the panel was an "invisible" dirigible. This was to be a giant, almost flat-shaped airship with a blue-tinted, nonreflective coating; it would cruise at an altitude of 90,000 feet along the borders of the Soviet Union at very slow speeds while using a large lens to photograph targets of interest.[31]

After completing these briefings at the end of February 1952, the BEACON HILL Study Group returned to MIT, where the panel members spent the next three months writing a report detailing their recommendations for ways to improve the amount and quality of intelligence being gathered on the Soviet Bloc. Published as a classified

[30] USAF, Project LINCOLN, *BEACON HILL Report: Problems of Air Force Intelligence and Reconnaissance*, Massachusetts Institute of Technology, 15 June 1952, pp. v, xi; app. A (~~S~~, downgraded to ~~C~~).

[31] Allen F. Donovan, telephone conversation with Donald E. Welzenbach, 21 June 1985 (U); James G. Baker, interview by Donald E. Welzenbach, tape recording, Washington, DC, 24 April 1985 (~~S~~).

document on 15 June 1952, the *BEACON HILL Report* advocated radical approaches to obtain the information needed for national intelligence estimates. Its 14 chapters covered radar, radio, and photographic surveillance; examined the use of passive infrared and microwave reconnaissance; and discussed the development of advanced reconnaissance vehicles. One of the report's key recommendations called for the development of high-altitude reconnaissance aircraft:

> *We have reached a period in history when our peacetime knowledge of the capabilities, activities and dispositions of a potentially hostile nation is such as to demand that we supplement it with the maximum amount of information obtainable through aerial reconnaissance. To avoid political involvements, such aerial reconnaissance must be conducted either from vehicles flying in friendly airspace, or—a decision on this point permitting—from vehicles whose performance is such that they can operate in Soviet airspace with greatly reduced chances of detection or interception.*[32]

Concern About the Danger of a Soviet Surprise Attack

The Air Force did not begin to implement the ideas of the *BEACON HILL Report* until the summer of 1953. By this time interest in reconnaissance had increased after Dwight D. Eisenhower became President in January 1953 and soon expressed his dissatisfaction with the quality of the intelligence estimates of Soviet strategic capabilities and the paucity of reconnaissance on the Soviet Bloc.[33]

To President Eisenhower and many other US political and military leaders, the Soviet Union was a dangerous opponent that appeared to be moving inexorably toward a position of military parity with the United States. Particularly alarming was Soviet progress in the area of nuclear weapons. In the late summer of 1949, the Soviet Union had detonated an atomic bomb nearly three years sooner than US experts had predicted. Then in August 1953—a scant nine months after the first US test of a hydrogen bomb—the Soviet Union detonated a hydrogen bomb manufactured from lithium deuteride, a technology more advanced than the heavy water method used by US

[32] *BEACON HILL Report*, pp. 164, 167-168 (~~S~~). This section of the report was written by Allen Donovan and Louis Ridenour.

[33] Lundahl and Brugioni interview (~~TS~~ Codeword).

scientists. Thus, new and extremely powerful weapons were coming into the hands of a government whose actions greatly disturbed the leaders of the West. Only two months before the successful hydrogen bomb test, Soviet troops had crushed an uprising in East Berlin. And, at the United Nations, the Soviet Bloc seemed bent on causing dissension between Western Europe and the United States and between the developed and undeveloped nations. This aggressive Soviet foreign policy, combined with advances in nuclear weapons, led officials such as Secretary of State John Foster Dulles to see the Soviet Union as a menace to peace and world order.

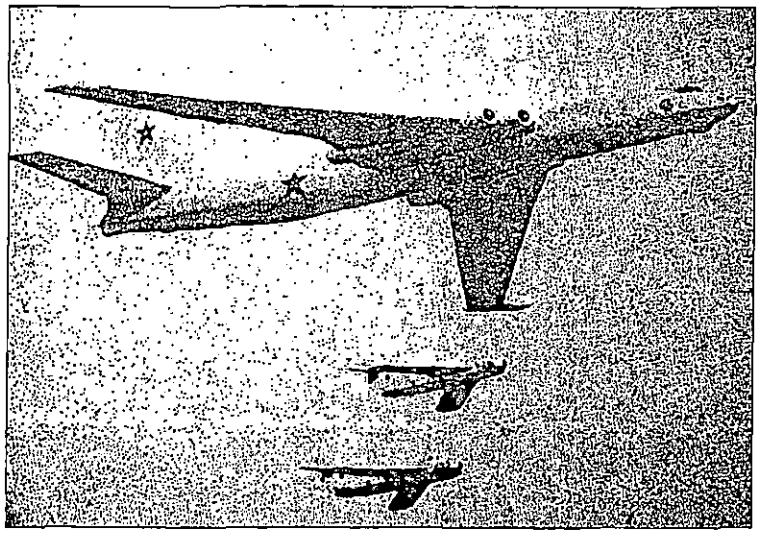
The Soviet Union's growing military strength soon became a threat not just to US forces overseas but to the continental United States itself. In the spring of 1953, a top secret RAND study pointed out the vulnerability of the SAC's US bases to a surprise attack by Soviet long-range bombers.[34]

Concern about the danger of a Soviet attack on the continental United States grew after an American military attache sighted a new Soviet intercontinental bomber at Ramenskoye airfield, south of Moscow, in 1953. The new bomber was the Myasishchev-4, later designated Bison by NATO. Powered by jet engines rather than the turboprops of Russia's other long-range bombers, the Bison appeared to be the Soviet equivalent of the US B-52, which was only then going into production. Pictures of the Bison taken at the Moscow May Day air show in 1954 had an enormous impact on the US intelligence community. Unlike several other Soviet postwar aircraft, the Bison was not a derivative of US or British designs but represented a native Soviet design capability that surprised US intelligence experts. This new long-range jet bomber, along with the Soviet Union's large numbers of older propeller and turboprop bombers, seemed to pose a significant threat to the United States, and, in the summer of 1954, newspapers and magazines began publishing articles highlighting the growing airpower of the Soviet Union. Pictures of the Bison bomber featured prominently in such stories.[35]

---

[34] RAND Corporation, Plans Analysis Section, "*Vulnerability of U.S. Strategic Power to a Surprise Attack in 1956*," RAND Special Memorandum No. 15, Santa Monica, California: the RAND Corporation, April 15, 1953 (TS, declassified May 1967).

[35] "AF Cites Red Bomber Progress," *Aviation Week*, May 24, 1954, p. 14; "Is Russia Winning the Arms Race?," *US News and World Report*, June 18, 1954, pp. 28-29; "Russia Parades Airpower as '*Big Stick*'," *Aviation Week*, June 28, 1954, p. 15; "Red Air Force: The World's Biggest," *Newsweek*, August 23, 1954, pp. 28-33.



Soviet Myasishchev-4 bomber (the Bison)

The Air Force Intelligence Systems Panel

Even before the publication of photographs of the Bison raised fears that the Soviet bomber force might eventually surpass that of the United States, the Air Force had already established a new advisory body to look for ways to implement the main recommendation of the BEACON HILL Report—the construction of high-flying aircraft and high-acuity cameras. Created in July 1953, the Intelligence Systems Panel (ISP) included several experts from the BEACON HILL Study Group: Land, Overhage, Donovan, and Miller. At the request of the Air Force, the CIA also participated in the panel, represented by Edward L. Allen of the Office of Research and Reports (ORR) and Philip Strong of the Office of Scientific Intelligence (OSI).[36]

The chairman of the new panel was Dr. James G. Baker, a research associate at the Harvard College Observatory. Baker had been involved in aerial reconnaissance since 1940, when he first advised the Army Air Corps on ways to improve its lenses. He then established a full-scale optical laboratory at Harvard—the Harvard University Optical Research Laboratory—to produce high-quality

[36] Memorandum for Robert Amory, Jr., Deputy Director, Intelligence from Edward L. Allen, Chief, Economic Research, ORR and Philip G. Strong, Chief, Operations Staff, OSI, "Meeting of the Intelligence Systems Panel of the Scientific Advisory Board, USAF," 26 August 1953, OSI (OSWR) records, job 80R-01424, box 1 ~~(S)~~.

lenses. Since the university did not wish to continue manufacturing cameras and lenses after the end of the war, the optical laboratory moved to Boston University, which agreed to sponsor the effort as long as the Air Force would fund it. Baker decided to remain at Harvard, so his assistant, Dr. Duncan E. Macdonald, became the new head of what was now called the Boston University Optical Research Laboratory (BUORL). Baker's association with the Air Force did not end with the transfer of the optical laboratory to Boston University, because he continued to design lenses to be used in photoreconnaissance.[37]

The ISP first met at Boston University on 3 August 1953. To provide background on the poor state of US knowledge of the Soviet Union, Philip Strong informed the other panel members that the best intelligence then available on the Soviet Union's interior was photography taken by the German Luftwaffe during World War II. Since the German photography covered only the Soviet Union west of the Urals, primarily west of the Volga River, many vital regions were not included. The ISP would, therefore, have to look for ways to provide up-to-date photography of all of the Soviet Union. Several Air Force agencies then briefed the panel members on the latest developments and proposed future projects in the area of aerial reconnaissance, including new cameras, reconnaissance balloons, and even satellites. Among the Air Force reconnaissance projects discussed were multiple sensors for use in existing aircraft such as the RB-47, RB-52, and RB-58; Project FICON—an acronym for "fighter conversion"—for adapting a giant, 10-engine B-36 bomber to enable it to launch and retrieve a Republic RF-84F Thunderflash reconnaissance aircraft; reconnaissance versions of the Navajo and Snark missiles; the high-altitude balloon program, which would be ready to go into operation by the summer of 1955; and the search for a new high-altitude reconnaissance aircraft.[38]

---

[37] Baker interview (S). In 1957, after the Air Force decided to cut back its funding of BUORL, Duncan Macdonald and Richard Leghorn (by then retired from the Air Force) formed their own corporation—Itek—and purchased the laboratory from Boston University (Leghorn interview ~~(S)~~).

[38] Memorandum for Robert Amory, Jr., Deputy Director, Intelligence, from Edward L. Allen, Chief, Economic Research, ORR, and Philip G. Strong, Chief, Operations Staff, OSI, "Meeting of the Intelligence Systems Panel of the Scientific Advisory Board, USAF," 26 August 1953; Memorandum for H. Marshall Chadwell, Assistant Director/Scientific Intelligence, from Chief, Support Staff, OSI, "Review of OSA Activities Concerned with Scientific and Technical Collection Techniques," 13 May 1955, p. 6, OSI (OSWR) records, job 80R-01424, box 1 (~~S~~); Donovan interview, 22 May 1985 ~~(S)~~.

The wide variety of programs discussed at the conference were all products of the Air Force's all-out effort to find a way to collect intelligence on the Communist Bloc. Some of the schemes went beyond the existing level of technology; others, like the camera-carrying balloons, were technically feasible but involved dangerous political consequences.

British Overflight of Kapustin Yar

The British were also working on high-altitude reconnaissance aircraft. In 1952 the Royal Air Force (RAF) began Project ROBIN, which was designed to modify the Canberra bomber for high-altitude reconnaissance. This project was probably inspired by Richard Leghorn's collaboration with English Electric Company designers in 1951, when they calculated ways to increase the altitude of the Canberra. The RAF equipped the new Canberra PR7 with Rolls-Royce Avon-109 engines and gave it long, fuel-filled wings. The range of this variant of the Canberra was now 4,300 miles, and, on 29 August 1955, it achieved an altitude of 65,880 feet.[39]

Sometime during the first half of 1953, the RAF employed a high-altitude Canberra on a daring overflight of the Soviet Union to photograph the missile test range at Kapustin Yar. Because of advanced warning from either radar or agents inside British intelligence, the overflight did not catch the Soviet Union by surprise. Soviet fighters damaged and nearly shot down the Canberra.[40] Rumors about this flight reached Washington during the summer of 1953, but official confirmation by the United Kingdom did not come until February 1954. While on a six-week tour of Europe to study aerial reconnaissance problems for the US Air Force's Scientific Advisory Board, James Baker was briefed by RAF intelligence officials on the Canberra overflight of the Soviet Union. On 22 and 23 March 1954, he reported on it to the full Scientific Advisory Board at Langley AFB, Virginia.

---

[39] Van der Aart, *Aerial Espionage*, p. 18; Philip G. Strong, Chief, Operations Staff, OSI, Memorandum for the Record, "Meeting of Air Force Scientific Advisory Board, 18-21 October 1953," 26 October 1953, OSI (OSWR) records, job 80R-01424, box 1 (~~TS~~, downgraded to ~~S~~).

[40] Stewart Alsop, *The Center*, (New York: Popular Library, 1968), p. 194; Beschloss, *Mayday*, pp. 78-79. Both of these books state that the project included the CIA, but there is no evidence to support this assertion.

Allen F. Donovan

Baker also chaired the next meeting of the Air Force's Intelligence Systems Panel in late April 1954 but could not tell its members about the British overflight of Kapustin Yar because they were not cleared for this information. The panel did, however, discuss the modifications for high-altitude flight being made to the US Canberra, the B-57.[41]

The Intelligence Systems Panel and the CL-282

The next Intelligence Systems Panel meeting took place on 24 and 25 May at Boston University and the Polaroid Corporation. Panel member Allen F. Donovan from the Cornell Aeronautical Laboratory evaluated the changes being made to the B-57 by the Martin Aircraft Company. Even without Martin's specifications or drawings, Donovan had been able to estimate what could be done to the B-57 by lengthening the wings and lightening the fuselage. He had determined that alterations to the B-57 airframe would not solve the reconnaissance needs expressed in the *BEACON HILL Report*. Theoretically, he explained to the panel, any multiengine aircraft built according to military specifications, including the B-57, would be too heavy to fly above 65,000 feet and hence would be vulnerable to Soviet interception. To be safe, Donovan explained, penetrating aircraft would need to fly above 70,000 feet for the entire mission.[42]

Development of such an aircraft was already under way, Donovan continued, for Philip Strong of the CIA had told him that the Lockheed Aircraft Corporation had designed a lightweight, high-flying aircraft. ISP chairman Baker then urged Donovan to travel to southern California to evaluate the Lockheed design and gather ideas on high-altitude aircraft from other aircraft manufacturers.

When he was finally able to make this trip in late summer, Donovan found the plane that he and the other ISP members had been seeking. On the afternoon of 2 August 1954, Donovan met with L. Eugene Root, an old Air Force acquaintance who was now a Lockheed vice-president, and learned about the Air Force's competition for a high-altitude reconnaissance aircraft. Kelly Johnson then showed Donovan the plans for Lockheed's unsuccessful entry. A lifelong sailplane enthusiast, Donovan immediately recognized that the

[41] Baker interview (~~S~~).

[42] Donovan interview (~~S~~); Baker interview (~~S~~).

CL-282 design was essentially a jet-propelled glider capable of attaining the altitudes that he felt were necessary to carry out reconnaissance of the Soviet Union successfully.[43]

Upon his return east on 8 August, Donovan got in touch with James Baker and suggested an urgent meeting of the Intelligence Systems Panel. Because of other commitments by the members, however, the panel did not meet to hear Donovan's report until 24 September 1954 at the Cornell Aeronautical Laboratory. Several members, including Land and Strong, were absent. Those who did attend were upset to learn that the Air Force had funded a closed competition for a tactical reconnaissance plane without informing them. But once Donovan began describing Kelly Johnson's rejected design for a jet-powered glider, they quickly forgot their annoyance and listened intently.

Donovan began by stressing that high-altitude reconnaissance aircraft had to fly above 70,000 feet to be safe from interception. Next, he set out what he considered to be the three essential requirements for a high-altitude spyplane: a single engine, a sailplane wing, and low structural load factors. Donovan strongly favored single-engine aircraft because they are both lighter and more reliable than multiengine aircraft. Although a twin-engine aircraft could theoretically return to base on only one engine, Donovan explained, it could only do so at a much lower altitude, about 34,000 feet, where it was sure to be shot down.

The second of Donovan's essential factors, a sailplane wing (in technical terms a high-aspect-ratio, low-induced-drag wing), was needed to take maximum advantage of the reduced thrust of a jet engine operating in the rarefied atmosphere of extreme altitude. Because of the thinness of the atmosphere above 70,000 feet, engineers estimated that the power curve of a jet engine would fall off to about 6 percent of its sea-level thrust.

Finally, low structural load factors, like those used by transport aircraft, were necessary to reduce weight and thereby achieve maximum altitude. Donovan explained that strengthening wings and

[43] Donovan interview (S)

wingroot areas to withstand the high speeds and sharp turns mandated by the standard military airworthiness rules added too much weight to the airframe, thereby negating the efficiency of the sailplane wing.

In short, it was possible to achieve altitudes in excess of 70, 000 feet, but only by making certain that all parts of the aeronautical equation were in balance: thrust, lift, and weight. The only plane meeting these requirements, Donovan insisted, was Kelly Johnson's CL-282 because it was essentially a sailplane. In Donovan's view, the CL-282 did not have to meet the specifications of a combat aircraft because it could fly safely above Soviet fighters.[44]

Donovan's arguments convinced the Intelligence Systems Panel of the merits of the CL-282 proposal, but this panel reported to the Air Force, which had already rejected the CL-282. Thus, even though the Lockheed CL-282 had several important sources of support by September 1954—the members of the Intelligence Systems Panel and high-ranking Air Force civilians such as Trevor Gardner—these backers were all connected with the Air Force. They could not offer funds to Lockheed to pursue the CL-282 concept because the Air Force was already committed to the Martin RB-57 and the Bell X-16. Additional support from outside the Air Force was needed to bring the CL-282 project to life, and this support would come from scientists serving on high-level advisory committees.

The Technological Capabilities Panel

The Eisenhower administration was growing increasingly concerned over the capability of the Soviet Union to launch a surprise attack on the United States. Early in 1954, Trevor Gardner had become alarmed by a RAND Corporation study warning that a Soviet surprise attack might destroy 85 percent of the SAC bomber force. Gardner then met with Dr. Lee DuBridge, President of the California Institute of Technology and Chairman of the Office of Defense Mobilization's Science Advisory Committee, and criticized the committee for not dealing with such essential problems as the possibility of a surprise attack. This criticism led DuBridge to invite Gardner to speak at the Science Advisory Committee's next meeting. After listening to

[44] Donovan interview (S); Baker interview (S).

Gardner, the committee members decided to approach President Eisenhower on the matter. On 27 March 1954, the President told them about the discovery of the Soviet Bison bombers and his concern that these new aircraft might be used in a surprise attack on the United States. Stressing the high priority he gave to reducing the risk of military surprise, the President asked the committee to advise him on this problem.[45]

James R. Killian, Jr.

The President's request led Chairman DuBridge to ask one of the most prominent members, MIT President James R. Killian, Jr., to meet with other Science Advisory Committee members in the Boston area to discuss the feasibility of a comprehensive scientific assessment of the nation's defenses. At their meeting at MIT on 15 April 1954, the group called for the recruitment of such a task force if the President endorsed the concept.

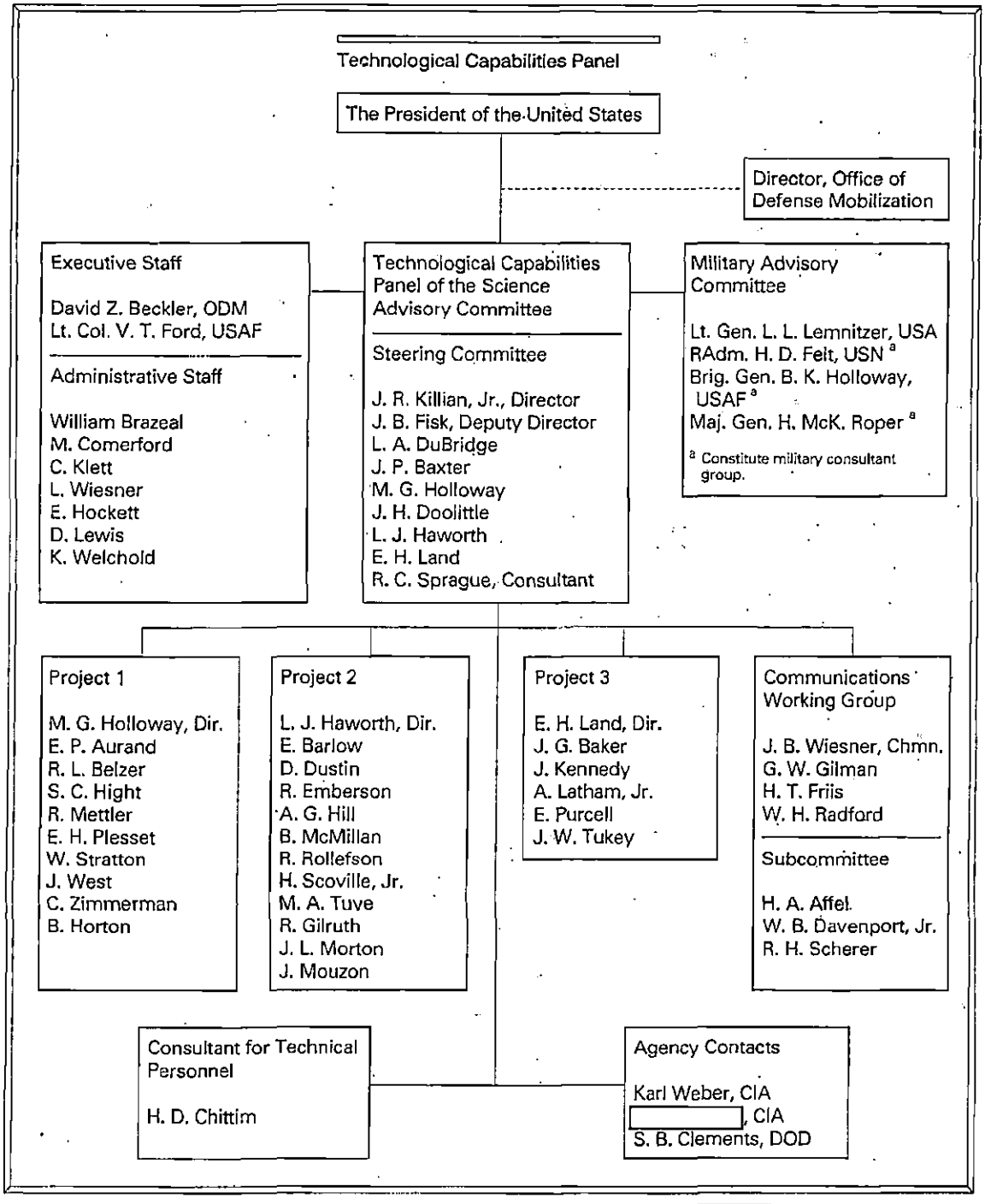
On 26 July 1954, President Eisenhower authorized Killian to recruit and lead a panel of experts to study "the country's technological capabilities to meet some of its current problems." Killian quickly set up shop in offices located in the Old Executive Office Building and organized 42 of the nation's leading scientists into three special project groups investigating US offensive, defensive, and intelligence capabilites, with an additional communications working group (see chart, page 28). The Technological Capabilities Panel (TCP) groups began meeting on 13 September 1954. For the next 20 weeks, the members of the various panels met on 307 separate occasions for briefings, field trips, conferences, and meetings with every major unit of the US defense and intelligence establishments. After receiving the most up-to-date information available on the nation's defense and intelligence programs, the panel members began drafting their report to the National Security Council.[46]

Project Three Support for the Lockheed CL-282

Even before the final Technological Capabilities Panel report was ready, one of the three working groups took actions that would have a major impact on the US reconnaissance program. Project Three had

[45] Beschloss, *Mayday*, pp. 73-74; Technological Capabilities Panel of the Science Advisory Committee, *Meeting the Threat of Surprise Attack*, 14 February 1955, p. 185 (hereafter cited as *TCP Report*) ~~(TS/~~Restricted Data, downgraded to ~~S)~~.

[46] James R. Killian, Jr., *Sputnik, Scientists, and Eisenhower: A Memoir of the First Special Assistant to the President for Science and Technology* (Cambridge: MIT Press, 1977), p. 68; Beschloss, *Mayday*, p. 74; *TCP Report*, pp. 185-186 ~~(S)~~.



Withheld under statutory authority of the Central Intelligence Agency Act of 1949 (50 U.S.C., section 403g)

the task of investigating the nation's intelligence capabilities. Its chairman was Edwin H. (Din) Land, the inventor of the polarizing filter and the instant camera. When James Killian asked Land to head Project Three, Land had to make a major decision about his career. At the time, the 45-year-old millionaire was on a leave of absence from Polaroid and was living in Hollywood, advising Alfred Hitchcock on the technological aspects of making three-dimensional movies. Land decided to give up his interest in cinema's third dimension and return east to Polaroid and the panel appointment.[47]

Edwin H. Land

Land's Project Three was the smallest of the three Technological Capabilities Panel projects, for he preferred what he called "taxicab committees"—committees small enough to fit into a single taxicab. The Project Three committee consisted of Land; James Baker and Edward Purcell of Harvard; chemist Joseph W. Kennedy of Washington University, St. Louis; mathematician John W. Tukey of Princeton University and Bell Telephone Laboratories; and Allen Latham, Jr., of Arthur D. Little, Inc., an engineer and former treasurer of the Polaroid Corporation.[48]

In mid-August 1954, Land and Baker went to Washington to arrange for the various intelligence organizations to brief the Project Three study group. As the briefings progressed, the panel members became more and more distressed at the poor state of the nation's intelligence resources. Land later noted, "We would go in and interview generals and admirals in charge of intelligence and come away worried. Here we were, five or six young men, asking questions that these high-ranking officers couldn't answer." Land added that the Project Three members were also not overly impressed with the Central Intelligence Agency.[49]

Land learned the details of Lockheed's proposed CL-282 aircraft soon after he arrived in Washington. Philip Strong showed him Kelly Johnson's conceptual drawing of the plane and told him that the Air Force had rejected it. Although Land had heard Allen Donovan

[47] James R. Killian, Jr., interview by Donald E. Welzenbach, tape recording, Cambridge, Massachusetts, 2 November 1984 (S); Land interview (~~TS~~ Codeword).

[48] *TCP Report*, p. 188 ~~(S)~~.

[49] Land interview (~~TS~~ Codeword).

briefly mention a Lockheed design for a high-flying aircraft at the 24-25 May meeting of Baker's Intelligence Systems Panel, he did not realize that that plane and the one in Strong's drawing were the same. As soon as Land saw Strong's copy of the CL-282 drawing, however, he telephoned Baker to say, "Jim, I think I have the plane you are after."[50]

A few days later, when Land showed Kelly Johnson's conceptual drawing to Baker and the other Project Three members, they all became enthusiastic about the aircraft's possibilities. Although Baker had heard Allen Donovan's brief mention of the Lockheed design in May, he had not yet seen a drawing of the aircraft because Donovan did not report to the ISP on his early-August trip to Lockheed until 24 September. After seeing the CL-282 drawing, Baker began designing a camera and lens system that would fit in the Lockheed craft.[51]

At the end of August, Land discussed the CL-282 with Allen Dulles's Special Assistant for Planning and Coordination, Richard Bissell, who came away from the meeting without any definite ideas as to what Land wanted to do with the aircraft. Overhead reconnaissance was not uppermost in Bissell's mind at the time, and it was unclear to him why he had even been contacted.[52] Bissell's outstanding academic credentials, his acquaintanceship with James Killian through his previous teaching experience at MIT, and his direct access to DCI Dulles may have led the Technological Capabilities Panel members to consider him the best CIA point of contact.

Although surprised that he had become involved in the CL-282 project, Bissell's interest was piqued, and he set out to learn what he could about reconnaissance systems. In early September 1954, Bissell had Douglas E. Ashford, a young Air Force officer on his staff, put together a general status report on air reconnaissance programs. Bissell forwarded the 16-page study to the Deputy Director of Central Intelligence (DDCI), Lt. Gen. Charles Pearre Cabell, USAF, on 24 September. In a covering memorandum, Bissell called Cabell's

[50] Baker interview (~~S~~).

[51] Ibid.

[52] Bissell interview (~~S~~).

attention to a section of the report about a "stripped or specialized aircraft" called the Lockheed CL-282.[53]

By September 1954, Land's Project Three study group had become very much interested in the Lockheed CL-282 design. Their interest grew even stronger when James Baker told them of Allen Donovan's strong case for the CL-282 at the 24-25 September meeting of the ISP. It is not possible to determine exactly when the Land committee decided to back the CL-282; in fact, there may never have been a formal decision as such. In view of Land's impulsive nature, he probably seized upon the CL-282 design as being a workable concept and immediately began developing it into a complete reconnaissance system.

During September and October the Project Three study group met frequently to discuss the Lockheed design and the reconnaissance equipment it would carry. Meetings were small, generally with fewer than 10 participants; Garrison Norton was often the only government official in attendance. At times outside experts joined in the proceedings. When the discussion turned to cameras and film, Land invited Dr. Henry Yutzy, Eastman Kodak's film expert, and Richard S. Perkin, President of the Perkin-Elmer Company, to participate. For discussions on the J57 engine, the panel members asked Perry W. Pratt, Pratt and Whitney's chief engineer, to attend. Kelly Johnson also met with the panel to review plans for the CL-282 system.[54]

By the end of October, the Project Three meetings had covered every aspect of the Lockheed design. The CL-282 was to be more than an airplane with a camera, it was to be an integrated intelligence-collection system that the Project Three members were confident could find and photograph the Soviet Union's Bison bomber fleet and, thus, resolve the growing "bomber gap" controversy. It was not just the Lockheed aircraft that had captured the Land group's fancy; the plane was seen as the platform for a whole new generation of aerial cameras that several committee members had been discussing since the BEACON HILL and Intelligence Systems Panel meetings. James Baker was in the process of developing a revolutionary new

[53] Memorandum for DDCI Charles Pearre Cabell from R. M. Bissell, Special Assistant to the Director for Planning and Coordination, "Aerial Reconnaissance," 24 September 1954, DCI Records, job 80-B-1676R, box 25 (~~TS~~, downgraded to S).

[54] Killian, *Sputnik, Scientists, and Eisenhower*, p. 82.

camera with tremendously improved resolution and film capacity, and the Eastman Kodak company was working on new thin, lightweight film.[55]

By October 1954, the Project Three study group had drafted a complete program for an overhead reconnaissance effort based on the CL-282 aircraft. The one remaining question was who would conduct the overflights. The committee's members, particularly Land, were not in favor of the Air Force conducting such missions in peacetime. Firmly believing that military overflights in armed aircraft could provoke a war, they argued for civilian overflights in unarmed, unmarked aircraft. In their view, the organization most suited for this mission was the Central Intelligence Agency.[56]

In late October 1954, the Project Three panel discussed the CL-282 system concept with DCI Allen Dulles and the Secretary of the Air Force's Special Assistant for Research and Development, Trevor Gardner. Dulles was reluctant to have the CIA undertake the project. He did not like to involve the CIA with military projects, even ones that the military had rejected, like the CL-282. Furthermore, the DCI strongly believed that the Agency's mission lay in the use of human operatives and secret communications, the classic forms of intelligence gathering. Land came away from this meeting with the impression that Dulles somehow thought overflights were not fair play. Project Three committee members were nevertheless convinced that technology, particularly in the form of the CL-282 and the new camera designs, would solve the nation's intelligence problems.[57]

A Meeting With the President

Allen Dulles's reluctance to involve the CIA in the CL-282 project did not stop the Project Three committee from pursuing its aims because it was able to go over Dulles's head and appeal directly to the President. Having participated in the BEACON HILL Study and the Intelligence Systems Panel, several Project Three members had definite ideas on how to improve intelligence collection, ideas that they were determined to present to the highest levels of government. They were able

[55] Land interview (~~TS~~ Codeword).

[56] Land interview (~~TS~~ Codeword); Baker interview (~~S~~).

[57] Land interview (~~TS~~ Codeword).

to do so because the Land committee was part of a panel commissioned by President Eisenhower to examine the nation's intelligence community and recommend changes. The committee thus had a direct line to the White House through James Killian's contacts there.

Early in November 1954, Land and Killian met with President Eisenhower to discuss high-altitude reconnaissance. Killian's memoirs contain an account of this crucial meeting:

> *Land described the [CL-282] system using an unarmed plane and recommended that its development be undertaken. After listening to our proposal and asking many hard questions, Eisenhower approved the development of the system, but he stipulated that it should be handled in an unconventional way so that it would not become entangled in the bureaucracy of the Defense Department or troubled by rivalries among the services.*[58]

The scientists from the advisory committees and the President were thus in agreement that the new reconnaissance program should be controlled by the CIA, not the military.

CIA and Air Force Agreement on the CL-282

Meanwhile Edwin Land and his Project Three colleagues were working to convince Allen Dulles that the CIA should run the proposed overflight program. On 5 November Land wrote to the DCI strongly urging that the CIA undertake the CL-282 project:

> *Here is the brief report from our panel telling why we think overflight is urgent and presently feasible. I [Land] am not sure that we have made it clear that we feel there are many reasons why this activity is appropriate for CIA, always with Air Force assistance. We told you that this seems to us the kind of action and technique that is right for the contemporary version of CIA: a modern and scientific way for an Agency that is always supposed to be looking, to do its looking. Quite strongly, we feel that you must always assert your first right to pioneer in scientific techniques for collecting intelligence—and choosing such partners to assist you as may be needed. This present opportunity for aerial photography seems to us a fine place to start.*[59]

---

[58] Killian, *Sputnik, Scientists, and Eisenhower*, p. 82. The exact date of the meeting cannot be determined, but it occurred during the first half of November 1954.

[59] Letter, Project Three Panel to DCI Allen F. Dulles, 5 November 1954, in *OSA History*, chap. 1, annex 1 (TS Codeword).

The letter had two attachments: a two-page summary of a complete operational plan for organizing, building, and deploying the CL-282 within a period of 20 months at a cost of $22 million and a three-page memorandum, entitled "A Unique Opportunity for Comprehensive Intelligence."

Aware of Dulles's preference for classical intelligence work, the Project Three memorandum stressed the superiority of the CL-282 program over traditional espionage methods:

> *We believe that these planes can go where we need to have them go efficiently and safely, and that no amount of fragmentary and indirect intelligence can be pieced together to be equivalent to such positive information as can thus be provided.*[60]

The Land committee memorandum also stressed the need for the CIA to undertake such reconnaissance missions rather than the Air Force, noting that "For the present it seems rather dangerous for one of our military arms to engage directly in extensive overflight." The committee members also listed the advantages of using the CL-282 rather than an Air Force aircraft:

> *The Lockheed super glider will fly at 70,000 feet, well out of the reach of present Russian interceptors and high enough to have a good chance of avoiding detection. The plane itself is so light (15,000 pounds), so obviously unarmed and devoid of military usefulness, that it would minimize affront to the Russians even if through some remote mischance it were detected and identified.*[61]

One additional advantage of the Lockheed design over the Air Force's proposed high-altitude reconnaissance aircraft was a faster completion time. Kelly Johnson had promised the Land committee that his aircraft would be flying by August 1955, just eight months after he proposed to start construction. The Bell X-16 prototype was not scheduled for completion before the spring of 1956.

The strong advocacy of Killian and the other scientists on the various advisory committees concerned with overhead reconnaissance, combined with President Eisenhower's support, finally won

---

[60] Memorandum for DCI Allen F. Dulles from Project Three Panel, "A Unique Opportunity for Comprehensive Intelligence," 5 November 1954, p. 3 (TS, downgraded to S) in *OSA History*, chap. 1, annex 1 (~~PS~~ Codeword).

[61] Ibid.

over DCI Dulles, but a project of this magnitude also required the support of the Air Force. Some Air Force officials, however, feared that a decision to build the CL-282 might jeopardize the Air Force's own RB-57 and X-16 projects. Just one month earlier, in October 1954, the Wright Air Development Command had appealed to the Air Force Deputy Chief of Staff for Development, Lt. Gen. Donald L. Putt, to oppose the adoption of the Lockheed design. The officials argued that the Bell X-16 was a better design because it was more airworthy than the CL-282 and could be used throughout the Air Force in different types of missions because it had two engines, wheels, and an armor-plated, pressurized pilot's compartment. If J57 engines were diverted to the CL-282, the appeal to General Putt warned, there would not be enough of these popular powerplants to meet the needs of the X-16 program.[62]

Having heard of the Wright Air Development Command attack on the CL-282, Allen Donovan of the Intelligence Systems Panel met with General Putt on 19 October to argue in favor of the Lockheed design. This discussion led General Putt to meet with 15 scientists from the Technological Capabilities Panel on 18 November 1954 to discuss the merits of the four proposed reconnaissance aircraft. Also present as a briefer was Maj. John Seaberg from the Wright Air Development Command, who later recalled:

> *What I did was present the results of my comparative analysis of all four designs. I showed the relative high altitude performance capabilities of all four. I pointed out that aerodynamically the Bell, Fairchild, and Lockheed designs were close. Martin's B-57, being a modification, was not quite as capable. I stated that, in my opinion, the J73 [General Electric engine] would not be good enough to do the job in Johnson's airplane. And further, I overlaid a curve showing that with the J57 [Pratt & Whitney engine] installed, it would then be competitive with the Bell and Fairchild designs.*[63]

This meeting—along with the knowledge that President Eisenhower also supported the CL-282—helped win over the Air Force. To be on the safe side, however, the Air Force did not abandon the X-16 program until the Lockheed aircraft had begun flying.

---

[62] Donovan interview (8).

[63] Quoted in Miller, *Lockheed U-2*, p. 13.

On 19 November, the day after Seaberg's briefing, the final decision on the CL-282 came at a luncheon hosted by Air Force Secretary Talbott. The participants—Dulles and Cabell from the CIA; Gardner, Ayer, and General Putt from the Air Force; Kelly Johnson; and Edwin Land—all agreed "that the special item of material described by Lockheed was practical and desirable and would be sought. . . . It was agreed that the Project should be a joint Air Force–CIA one but that, regardless of the source of the funds, whether AF or CIA, CIA unvouchered channels would be needed to pass the funds."[64]

It is interesting to note that Lockheed, which had originally developed the CL-282 on its own and had devoted considerable effort to promoting it, had to be persuaded to undertake the project in November 1954 because the company had become heavily committed to several other civilian and military projects. When Kelly Johnson received a call from Trevor Gardner on 17 November asking him to come to Washington for conversations on the project, his instructions from Lockheed's senior management were "to not commit to any program during the visit, but to get the information and return." When he returned to California, Johnson noted in his project log that "I was impressed with the secrecy aspect and was told by Gardner that I was essentially being drafted for the project. It seemed, in fact, that if I did not talk quietly, I might have to take a leave of absence from my job at Lockheed to do this special project."[65] Of course, Kelly Johnson did not need to be drafted or persuaded into undertaking such a bold step forward in aircraft design. He used Gardner's statement to convince Lockheed's senior management to approve the project, which they did after meeting with Johnson when he returned to California on the evening of 19 November.

Four days later, on 23 November, the Intelligence Advisory Committee (IAC) approved DCI Dulles's request to undertake the CL-282 project. The following day Dulles signed a three-page memorandum, drafted by DDCI Cabell, asking President Eisenhower to approve the overhead reconnaissance project. That same afternoon, at a meeting attended by the Secretaries of State and Defense and senior Air Force officials, Dulles and Cabell presented the document to the

[64] Charles Pearre Cabell, Memorandum for the Record, "Luncheon Meeting with the Secretary of the Air Force," 19 November 1954, in *OSA History*, chap. 2, annex 4 (~~PS~~ Codeword).

[65] Johnson, "Log for Project X," 17 and 19 November 1954.

President and received verbal authorization to proceed. Eisenhower told Dulles that the project was to be managed by the Agency and that the Air Force was to provide any assistance needed to get it operational.[66]

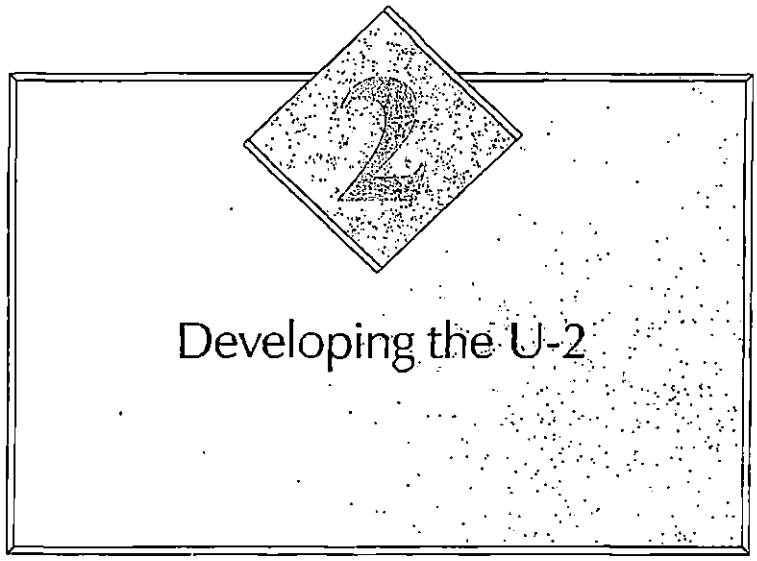
Thus, it was that the CIA entered into the world of high technology primarily because of decisions and actions taken outside the Agency: the Air Force's refusal to build the CL-282 aircraft, President Eisenhower's desire to have a sensitive overflight project conducted by a civilian agency rather than the military, and, above all, the determination by a small group of prominent scientists that the Lockheed design represented the best possible overhead reconnaissance system.[67]

---

[66] Charles Pearre Cabell, Memorandum for the Record, "Meeting at the White House," 24 November 1954, in *OSA History*, chap. 2, annex 8 (TS Codeword) ; Beschloss, *Mayday*, pp. 82-83; Andrew J. Goodpaster, Memorandum of Conference with the President, 24 November 1954," White House Office of the Staff Secretary, Alpha Series, Dwight D. Eisenhower Library (hereafter cited as WHOSS, Alpha, DDEL) (TS, declassified).

[67] Scientists remained active in advising the government on overhead reconnaissance. In February 1955, the Technological Capabilities Panel issued its final report, which strongly urged the use of technology to gather intelligence. President Eisenhower strongly backed the panel's findings and directed government agencies to respond to the recommendations by June. The CIA's most important reaction to the Technological Capabilities Panel report was to create its own Scientific Advisory Board composed of the members of the Project Three Study Group with the addition of James Killian and Jerome B. Wiesner, professor of electrical engineering at MIT. Edwin Land served as chairman of the CIA Scientific Advisory Board for the next 10 years, and it soon became known unofficially as the Land Panel. This panel provided important advice to the Agency, particularly in the field of overhead reconnaissance.

President Eisenhower also acted to increase the amount and quality of scientific advice he was receiving. In January 1956 he established the President's Board of Consultants on Foreign Intelligence Activities (renamed the President's Foreign Intelligence Advisory Board in 1961) to oversee the intelligence community and advise him on intelligence matters. The board's first chairman was James Killian. In 1957 the President reorganized and upgraded the Office of Defense Mobilization's Science Advisory Committee, which became the President's Science Advisory Committee. He also named James Killian to be the first Special Assistant to the President for Science and Technology. In this new position, Killian served as the President's scientific advisor and the chairman of the President's Scientific Advisory Committee (Killian stepped down as chairman of the President's Board of Consultants on Foreign Intelligence Activities but remained a member). These actions by the President brought scientists into the White House and gave them considerable influence.



THE ESTABLISHMENT OF THE U-2 PROJECT

On 26 November 1954, the day after Thanksgiving, Allen Dulles called his special assistant, Richard Bissell, into his office to tell him that President Eisenhower had just approved a very secret program and that Dulles wanted Bissell to take charge of it. Saying it was too secret for him to explain, Dulles gave Bissell a packet of documents and told him he could keep it for several days to acquaint himself with the project. Bissell had long known of the proposal to build a high-altitude reconnaissance aircraft, but only in the most general terms. Now he learned in detail about the project that proposed sending aircraft over the Soviet Union.

Late on the morning of 2 December 1954, Dulles told Bissell to go to the Pentagon on the following day to represent the Agency at an organizational meeting for the U-2[1] project. Before leaving, Bissell asked Dulles which agency was to run the project. The DCI replied that nothing had been clearly decided. Bissell then asked who was going to pay for the project. Dulles answered: "That wasn't even mentioned. You'll have to work that out." [2]

Bissell was accompanied by Herbert I. Miller, chief of the Office of Scientific Intelligence's Nuclear Energy Division, who soon became the executive officer of the overflight project. When Bissell and Miller arrived at the Pentagon on the afternoon of 3 December, they

[1] Although the Lockheed CL-282 was not designated as the U-2 until July 1955, this study will use the more widely known designator to avoid confusion.

[2] Bissell interview (~~S~~); *OSA History*, chap. 3, p. 1 (TS Codeword).

James A. Cunningham, Jr.

sat down with a group of key Air Force officials that included Trevor Gardner and Lt. Gen. Donald L. Putt. The participants spent very little time delineating Air Force and Agency responsibilities in the project, taking for granted that the CIA would handle the security matters. Much of the discussion centered on methods for diverting Air Force materiel to the program, particularly the Pratt & Whitney J57 engines, because a separate contract for the engines might jeopardize the project's security. The Air Force promised to turn over a number of J57 engines, which were then being produced for B-52s, KC-135s, F-100s, and RB-57s. Eventually Bissell asked who was going to pay for the airframes to be built by Lockheed. His query was greeted with silence. Everyone present had their eyes on him because they all expected the Agency to come up with the funds. Bissell rose from his chair, said he would see what he could do, and the meeting adjourned.[3]

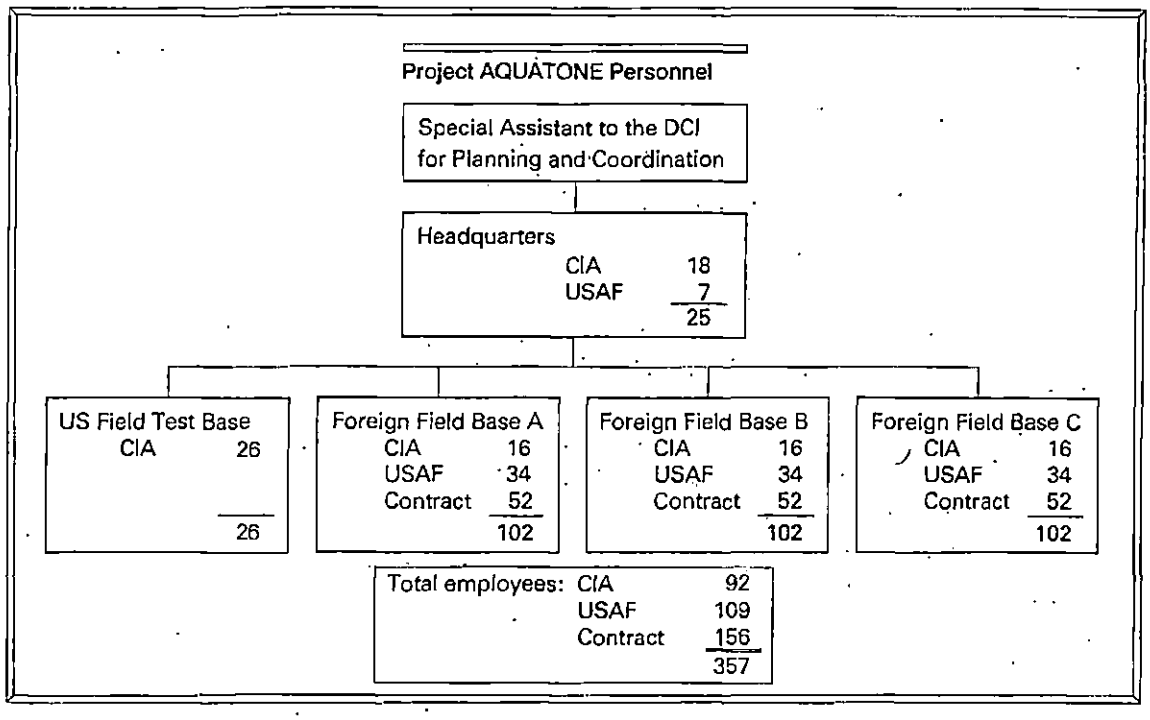
After the meeting, Bissell told Dulles that the CIA would have to use money from the Contingency Reserve Fund to get the project going. The DCI used this fund to pay for covert activities, following approval by the President and the Director of the Budget. Dulles told Bissell to draft a memorandum for the President on funding the overflight program and to start putting together a staff for Project AQUATONE, the project's new codename.

At first the new "Project Staff" (renamed the Development Projects Staff in April 1958) consisted of Bissell, Miller, and the small existing staff in Bissell's Office of the Special Assistant to the DCI. During the months that followed the establishment of the project, its administrative workload increased rapidly, and in May 1955 the project staff added an administrative officer, James A. Cunningham, Jr., a former Marine Corps pilot then working in the Directorate of Support. Cunningham stayed with the U-2 project for the next 10 years. Two other key project officials who began their duties early in 1955 were [redacted] the finance officer, and George Kucera, the contracting officer.[4]

Withheld under statutory authority of the Central Intelligence Agency Act of 1949 (50 U.S.C., section 403g)

[3] *OSA History*, chap. 3, p. 2 (TS Codeword); Bissell interview, 8 November 1984 ~~(S)~~; Beschloss, *Mayday*, p. 89.

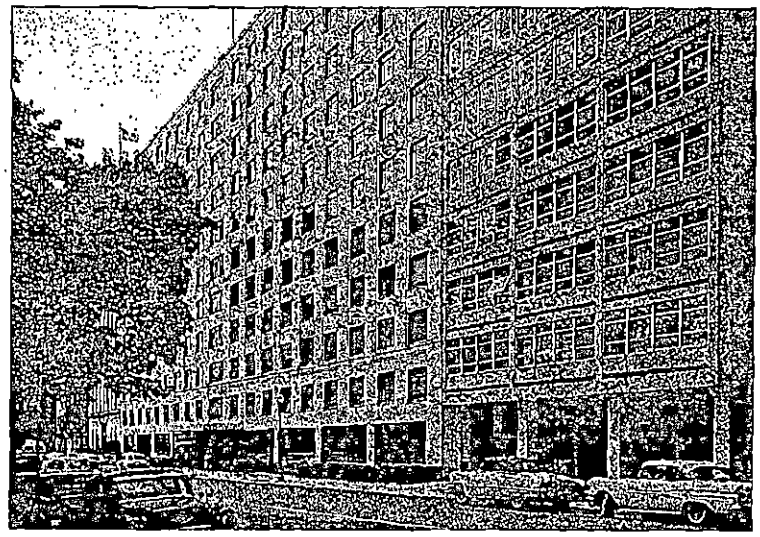
[4] *OSA History*, chap. 3, pp. 6-7, chap. 4, pp. 1-2, chap. 5, pp. 27-29 (TS Codeword); *Chronology of the Office of Special Activities, 1954-1968*, (CIA: DS&T, 1969), p. 2-4 ~~(TS~~ Codeword) (hereafter cited as *OSA Chronology*).



During the first half of 1955, the project staff grew slowly; many of the individuals working on overhead reconnaissance remained on the rolls of other Agency components. To achieve maximum security, Bissell made the project staff self-sufficient. Project AQUATONE had its own contract management, administrative, financial, logistic, communications, and security personnel, and, thus, did not need to turn to the Agency directorates for assistance. Funding for Project AQUATONE was also kept separate from other Agency components; its personnel and operating costs were not paid out of regular Agency accounts. As approving officer for the project, Richard Bissell could obligate funds in amounts up to $100,000; larger sums required the DCI's approval.[5]

At the end of April 1955, Bissell's staff developed, and the Deputy Director for Support approved, the first table of organization for Project AQUATONE. Once operational, the project would have a

[5] *OSA History*, chap. 3, pp. 5-7 (TK Codeword).



The Matomic Building

total of 357 personnel divided among project headquarters, a US testing facility, and three foreign field bases. CIA employees represented only one-fourth (92) of the total. The Air Force personnel commitment was larger, with 109 positions on the 1955 table of organization (this total does not include many other Air Force personnel, such as SAC meteorologists, who supported the U-2 project in addition to their other duties). The largest Project AQUATONE category was contract employees, with 156 positions in 1955. This category included maintenance and support personnel from Lockheed (five per aircraft), the pilots, and support personnel from other contractors for items such as photographic equipment.[6]

The first project headquarters was in CIA's Administration (East) Building at 2430 E Street, NW. Continued growth caused the AQUATONE staff to move several times during its first two years. On 1 May 1955, the project staff moved to the third floor of a small red brick building (the Briggs School) at 2210 E Street, NW. Then on 3 October, the staff moved to Wings A and C of Quarters Eye, a World War II "temporary" building on Ohio Drive, NW, in the West Potomac Park area of Washington. On 25 February 1956, the project staff moved again, this time to the fifth floor of the Matomic Building

[6] Project AQUATONE Table of Organization, 28 April 1955 in *OSA History*, chap. 3, annex 15 (~~TS~~ Codeword).

at 1717 H Street, NW. Here the staff remained for the next six years until it moved into the new CIA Headquarters building at Langley in March 1962. The final move came in January 1968, when the project staff (by that time known as the Office of Special Activities) moved to the Tyler Building in Tysons Corner, Virginia.[7]

Bissell reported directly to the DCI, although in reality the DDCI, Gen. Charles Pearre Cabell, was much more closely involved in the day-to-day affairs of the overhead reconnaissance project. Cabell's extensive background in Air Force intelligence, particularly in overhead reconnaissance, made him ideally qualified to oversee the U-2 project. Cabell frequently attended White House meetings on the U-2 for the DCI.

DDCI Charles Pearre Cabell

FUNDING ARRANGEMENTS FOR PROJECT AQUATONE

Although Allen Dulles had approved the concept of covert funding for the reconnaissance project, many financial details remained to be settled, including the contract with Lockheed. Nevertheless, work on the U-2 began as soon as the project was authorized. Between 29 November and 3 December 1954, Kelly Johnson pulled together a team of 25 engineers, which was not easy because he had to take them off other Lockheed projects without being able to explain why to their former supervisors. The engineers immediately began to work 45 hours a week on the project. The project staff gradually expanded to a total of 81 personnel, and the workweek soon increased to 65 hours.[8]

Kelly Johnson's willingness to begin work on the aircraft without a contract illustrates one of the most important aspects of this program: the use of unvouchered funds for covert procurement. Lockheed was well acquainted with the covert procurement process, having previously modified several aircraft for covert use by the CIA.

Covert funding for sensitive projects simplifies both procurement and security procedures because the funds are not attributable to the Federal Government and there is no public accountability for their

[7] *OSA History*, chap. 18, pp. 7-8 (TS Codeword); *OSA Chronology*, pp. 4, 7, 10, 45 (TS Codeword).

[8] Johnson, "Log for Project X," 29 November–3 December 1954 (U).

use. Public Law 110, approved by the 81st Congress on 20 June 1949, designates the Director of Central Intelligence as the only government employee who can obligate Federal money without the use of vouchers. By using unvouchered funds, it is possible to eliminate competitive bidding and thereby limit the number of parties who know about a given project. The use of unvouchered funds also speeds up the Federal procurement cycle. A general contractor such as Lockheed can purchase much, if not all, of the supplies needed for a project without resorting at each step to the mandated procurement procedures involving public, competitive bidding.

25X1, E.O.13526

In mid-December 1954, President Eisenhower authorized DCI Dulles to use $35 million from the Agency's Contingency Reserve Fund to finance the U-2 project. Then on 22 December 1954, the Agency signed a letter contract with Lockheed, using the codename Project OARFISH. The Agency had proposed to give Lockheed "performance specifications" rather than the standard Air Force "technical specifications," which were more rigid and demanding, and Kelly Johnson agreed that such a move would save a lot of money. Lockheed's original proposal to the Air Force in May 1954 had been $28 million for 20 U-2s equipped with GE J73 engines. During negotiations with CIA General Counsel Lawrence R. Houston, Lockheed changed its proposal to $26 million for 20 airframes plus a two-seat trainer model and spares; the Air Force was to furnish the engines. Houston insisted that the Agency could only budget $22.5 million for the airframes because it needed the balance of the available $35 million for cameras and life-support gear. The two sides finally agreed on a fixed-price contract with a provision for a review three-fourths of the way through to determine if the costs were going to exceed the $22.5 million figure. The formal contract, No. SP-1913, was signed on 2 March 1955 and called for the delivery of the first U-2 in July 1955 and the last in November 1956. Meanwhile, to keep work moving at Lockheed, Richard Bissell wrote a check for $1,256,000 and mailed it to Kelly Johnson's home on 21 February 1955.[9]

25X1, E.O.13526

[9] John S. Warner, Office of the General Counsel, interview by Donald E. Welzenbach, Washington, DC, tape recording, 5 Aug 1983 (S); *OSA History*, chap. 5, pp. 1-2 and annex 42 (TS Codeword); Johnson, "Log for Project X," 21 February 1955.

As it turned out, no review of the contract was necessary at the three-fourths point. Lockheed delivered the aircraft not only on time but under budget. During the final contract negotiations in the spring of 1958, Lockheed and the US Government agreed on a price for the original 20 aircraft of $17,025,542 plus a profit of $1,952,055 for a total of $18,977,597—less than $1 million for each aircraft. Because its design was based on Lockheed's F-104, the U-2 was relatively inexpensive even though only a small number of aircraft had been ordered. Only the wings and tail were unique; Lockheed manufactured the other portions of the aircraft using the F-104's jigs and dies.

MAJOR DESIGN FEATURES OF THE U-2

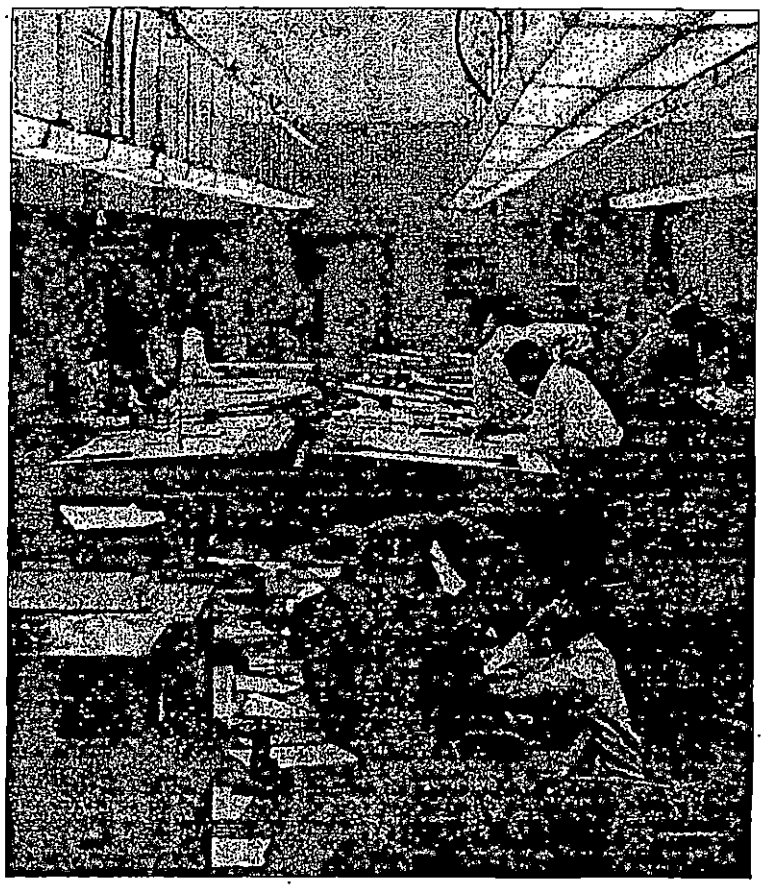
Aware of the great need for secrecy in the new project, Kelly Johnson placed it in Lockheed's Advanced Development facility at Burbank, California, known as the Skunk Works.[10] Lockheed had established this highly secure area in 1945 to develop the nation's first jet aircraft, the P-80 Shooting Star. The small Skunk Works staff began making the detailed drawings for the U-2, which was nicknamed the "Angel" because it was to fly so high.

Kelly Johnson's approach to prototype development was to have his engineers and draftsmen located not more than 50 feet from the aircraft assembly line. Difficulties in construction were immediately brought to the attention of the engineers, who gathered the mechanics around the drafting tables to discuss ways to overcome the difficulties. As a result, engineers were generally able to fix problems in the design in a matter of hours, not days or weeks. There was no emphasis placed on producing neatly typed memorandums; engineers simply made pencil notations on the engineering drawings in order to keep the project moving quickly.[11]

A little more than a week after he had been authorized to begin the project, Kelly Johnson wrote a 23-page report detailing his most recent ideas on the U-2 proposal. The aircraft, he explained, would be designed to meet load factors of only 2.5 g's, which was the limit for transport aircraft rather than combat planes. The U-2 would have a

[10] The Lockheed "Skunk Works" was named after the Kickapoo Joy Juice factory known as the "Skonk Works" in Al Capp's comic strip *Li'l Abner*.

[11] Ben A. Rich (current head of the "Skunk Works"), interview by Donald E. Welzenbach and Gregory W. Pedlow, Burbank, California, 26 August 1988.



Skunk Works Design Staff

speed of Mach 0.8 or 460 knots at altitude. Its initial maximum altitude would be 70,600 feet and the ultimate maximum altitude would be 73,100 feet. According to these early December 1954 specifications, the new plane would take off at 90 knots, land at 76 knots, and be able to glide 244 nautical miles from an altitude of 70,000 feet. After discussing the reconnaissance bay with James Baker, Johnson had worked out various equipment combinations that would not exceed the weight limit of 450 pounds. Johnson ended his report by promising the first test flight by 2 August 1955 and the completion of four aircraft by 1 December 1955.[12]

[12] Kelly Johnson, "A High-Altitude Reconnaissance Aircraft," 9 December 1954, Lockheed Contract Files, OSA Records (S).

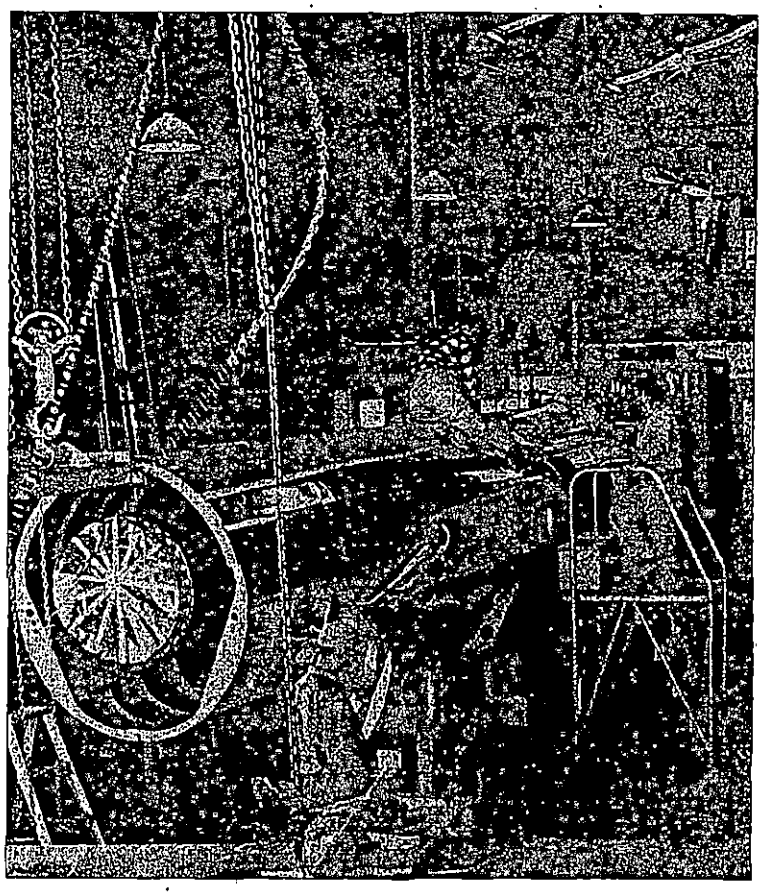
In designing the U-2 aircraft, Kelly Johnson was confronted with two major problems—fuel capacity and weight. To achieve intercontinental range, the aircraft had to carry a large supply of fuel, yet, it also had to be light enough to attain the ultrahigh altitudes needed to be safe from interception. Although the final product resembled a typical jet aircraft, its construction was unlike any other US military aircraft. One unusual design feature was the tail assembly, which—to save weight—was attached to the main body with just three tension bolts. This feature had been adapted from sailplane designs.

The wings were also unique. Unlike conventional aircraft, whose main wing spar passes through the fuselage to give the wings continuity and strength, the U-2 had two separate wing panels, which were attached to the fuselage sides with tension bolts (again, just as in sailplanes). Because the wing spar did not pass through the fuselage, Johnson was able to locate the camera behind the pilot and ahead of the engine, thereby improving the aircraft's center of gravity and reducing its weight.

The wings were the most challenging design feature of the entire airplane. Their combination of high-aspect ratio and low-drag ratio (in other words, the wings were long, narrow, and thin) made them unique in jet aircraft design. The wings were actually integral fuel tanks that carried almost all of the U-2's fuel supply.

The fragility of the wings and tail section, which were only bolted to the fuselage, forced Kelly Johnson to look for a way to protect the aircraft from gusts of wind at altitudes below 35,000 feet, which otherwise might cause the aircraft to disintegrate. Johnson again borrowed from sailplane designs to devise a "gust control" mechanism that set the ailerons and horizontal stabilizers into a position that kept the aircraft in a slightly nose-up attitude, thereby avoiding sudden stresses caused by wind gusts. Nevertheless, the U-2 remained a very fragile aircraft that required great skill and concentration from its pilots.

The final major design feature was the lightweight, bicycle-type landing gear. The entire structure—a single oleostrut with two lightweight wheels toward the front of the aircraft and two small, solid-mount wheels under the tail—weighed only 208 pounds yet could withstand the force of touchdown for this 7-ton aircraft. Because both sets of wheels were located underneath the fuselage, the U-2 was also equipped with detachable pogos (long, curved sticks with two small wheels on them) on each wing to keep the wings level during takeoff. The pilot would drop the pogos immediately after takeoff so



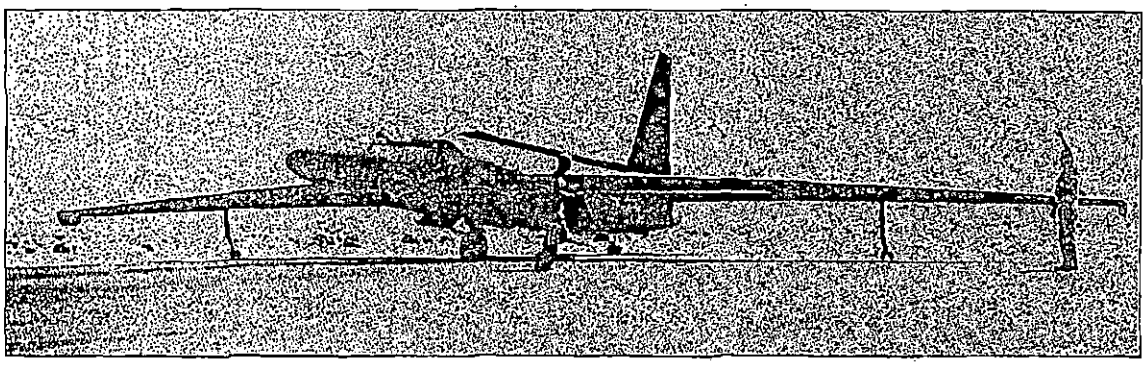
U-2 at testing site before attachment of wings and tail assembly

that they could be recovered and reused. The aircraft landed on its front and back landing gear and then gradually tilted over onto one of the wingtips, which were equipped with landing skids.[13]

THE DEVELOPMENT OF THE CAMERA SYSTEM

By December 1954, Kelly Johnson was at work on drawings for the U-2's airframe and Pratt & Whitney was already building the J57 jet

[13] For the design features of the U-2 in early 1955, see R. F. Boehme, *Summary Report: Reconnaissance Aircraft*, Lockheed Aircraft Corporation Report 10420, 28 January 1955, pp. 7-9, OSA Records, job 74-B-645, box 1 (S).



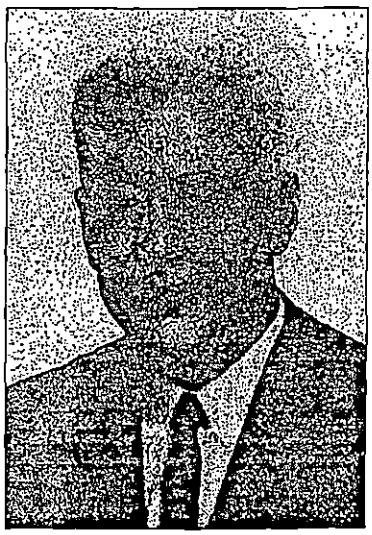
U-2 landing gear and pogos

engine, but no firm plans existed for the all-important cameras. Existing cameras were too bulky and lacked sufficient resolution to be used in high-altitude reconnaissance.

The workhorses of World War II aerial photography had been the Fairchild K-19 and K-21 framing cameras with lenses of varying focal lengths from 24 to 40 inches. Late in the war, the trimetrogon K-17 mapping-camera system came into use. This system consisted of three separate cameras which made three photographs simultaneously: a vertical, an oblique to the left, and an oblique to the right. The major shortcomings of the trimetrogon system were the large amount of film required and the system's lack of sharp definition on the obliques.

The standard aerial cameras available in the early 1950s could achieve resolutions of about 20 to 25 feet (7 to 8 meters) on a side when used at an altitude of 33,000 feet (10,000 meters), or about 25 lines per millimeter in current terms of reference. Such resolution was considered adequate because aerial photography was then used primarily to choose targets for strategic bombing, to assess bomb damage after air raids, and to make maps and charts. Unfortunately, a camera with a resolution of only 20 to 25 feet at a height of 33,000 feet was too crude to be used at twice that altitude. Indeed, for intelligence purposes a resolution of less than 10 feet was necessary to discern smaller targets in greater detail. This meant that any camera carried to altitudes above 68, 000 feet had to be almost four times as good as existing aerial cameras in order to achieve a resolution of less than 10 feet. As a result, some scientists doubted that useful photography could be obtained from altitudes higher than 40,000 feet.[14]

[14] Baker interview (S).



James G. Baker

The first success in designing very-high-acuity lenses came in the mid-1940s, when James G. Baker of Harvard and Richard S. Perkin of the Perkin-Elmer (P-E) Company of Norwalk, Connecticut, collaborated on a design for an experimental camera for the Army Air Force. They developed a 48-inch focal-length scanning camera that was mounted in a modified B-36 bomber. When tested over Fort Worth, Texas, at 34,000 feet, the new camera produced photographs in which two golf balls on a putting green could be distinguished (in reality, however, the "golf balls" were 3 inches in diameter). These photographs demonstrated the high acuity of Baker's lens, but the camera weighed more than a ton and was much too large to be carried aloft in an aircraft as small as the U-2.

Realizing that size and weight were the major restraining factors in developing a camera for the U-2, James Baker began working on a radically new system in October 1954, even before the CIA adopted the Lockheed proposal. Baker quickly recognized, however, that he would need almost a year to produce a working model of such a complex camera. Since Kelly Johnson had promised to have a U-2 in the air within eight months, Baker needed to find an existing camera that could be used until the new camera was ready. After consulting with his friend and colleague Richard Perkin, Baker decided to adapt for the U-2 an Air Force camera known as the K-38, a 24-inch aerial framing camera built by the Hycon Manufacturing Company of Pasadena, California.

Perkin suggested modifying several standard K-38 cameras in order to reduce their weight to the U-2's 450-pound payload limit. At the same time, Baker would make critical adjustments to existing K-38 lenses to improve their acuity. Baker was able to do this in a few weeks, so several modified K-38s, now known as A-1 cameras, were ready when the first "Angel" aircraft took to the air in mid-1955.[15]

CIA awarded Hycon a contract for the modified K-38 cameras, and Hycon, in turn, subcontracted to Perkin-Elmer to provide new lenses and to make other modifications to the cameras in order to make them less bulky. In its turn, Perkin-Elmer subcontracted to Baker to rework the existing K-38 lenses and later design an improved lens system. To keep his lens-designing efforts separate from

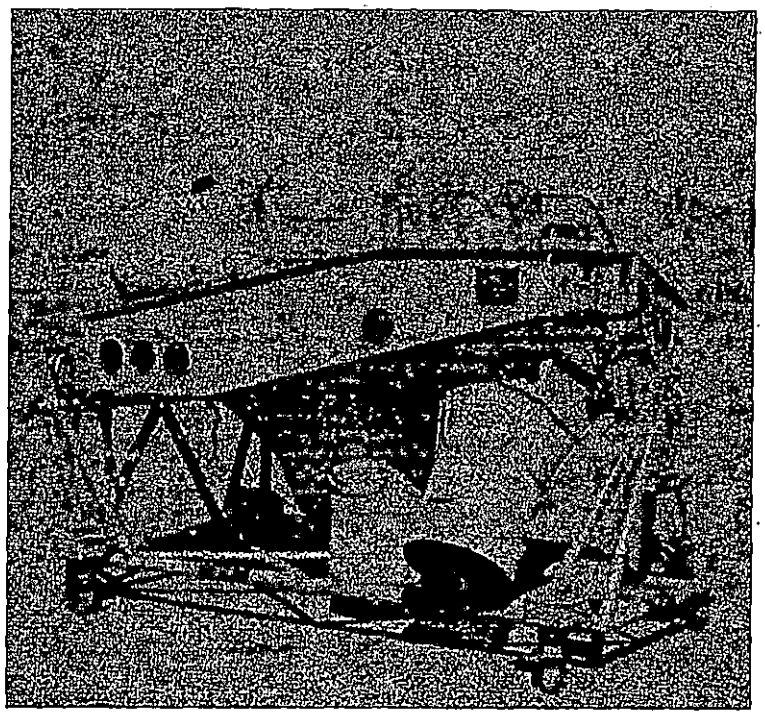
[15] *Ibid.*

A-1 camera

his research associate duties at Harvard and his service on government advisory bodies, Baker established a small firm known as Spica, Incorporated, on 31 January 1955.

The A-1 camera system consisted of two 24-inch K-38 framing cameras. One was mounted vertically and photographed a 17.2° swath beneath the aircraft onto a roll of 9.5-inch film. The second K-38 was placed in a rocking mount so that it alternately photographed the left oblique and right oblique out to 36.5° onto separate rolls of 9.5-inch film. The film supplies unwound in opposite directions in order to minimize their effect on the balance of the aircraft. Both cameras used standard Air Force 24-inch focal-length lenses adjusted for maximum acuity by Baker. The development of the special rocking mount by Perkin-Elmer's Dr. Roderic M. Scott was a major factor in reducing the size and weight of the A-1 system, because the mount provided broad transverse coverage with a single lens, ending the need for two separate cameras.[16]

[16] *OSA History*, chap. 1, annex 3, pp. 1-3 (~~TS~~ Codeword).



A-2 camera

U-2s equipped with the A-1 camera system also carried a Perkin-Elmer tracking camera using 2.75-inch film and a 3-inch lens. This device made continuous horizon-to-horizon photographs of the terrain passing beneath the aircraft. Because the A-1 system was new, it also included a backup camera system, a K-17 6-inch three-camera trimetrogon unit using 9-inch film.

While the A-1 system was still being developed, James Baker was already working on the next generation of lenses for high-altitude reconnaissance. Baker was a pioneer in using computers to synthesize optical systems. His software algorithms made it possible to model lens designs and determine in advance the effects that variations in lens curvatures, glass compounds, and lens spacings would have on rays of light passing through a lens. These "ray-tracing" programs required extensive computations, and, for this he turned to the most modern computer available, an IBM CPC (card-programmed calculator) installation at nearby Boston University.[17]

[17] Ibid., chap. 1, pp. 7-8 (~~TS~~ Codeword).

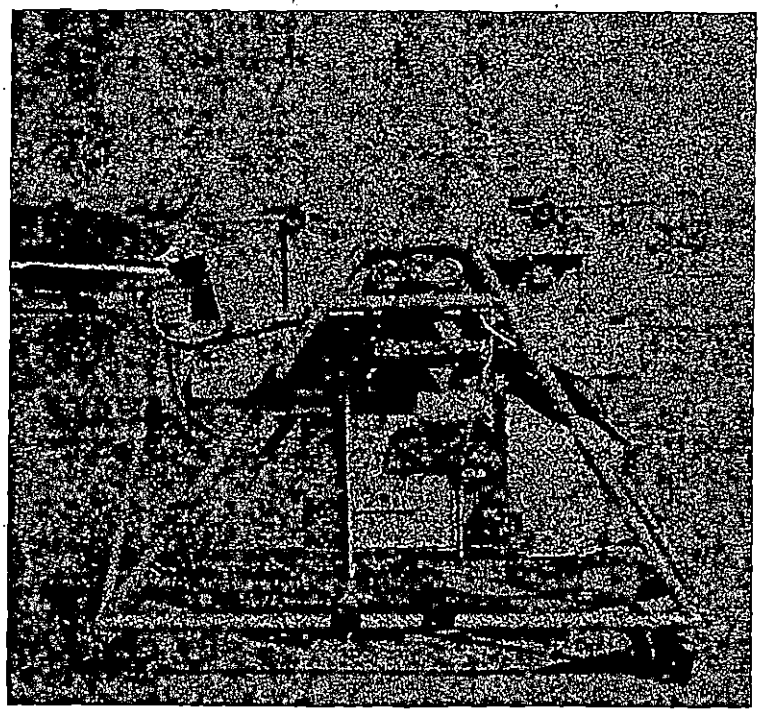
Baker's new lenses were used in a camera system known as the A-2, which returned to a trimetrogon arrangement because of problems with the A-1 system's rocking mount. The A-2 consisted of three separate K-38 framing cameras and 9.5-inch film magazines. One K-38 filmed the right oblique, another the vertical, and a third the left oblique. The A-2 system also included a 3-inch tracking camera. All A-2 cameras were equipped with the new 24-inch f/8.0 Baker-designed lenses. These were the first relatively large photographic objective lenses to employ several aspheric surfaces. James Baker personally ground these surfaces and made the final bench tests on each lens before releasing it to the Agency. These lenses were able to resolve 60 lines per millimeter, a 240-percent improvement over existing lenses.[18]

Once Baker and Scott had redesigned the 24-inch lens for the K-38 devices, they turned their attention to Baker's new camera design, known as the B model. It was a totally new concept, a high-resolution panoramic-type framing camera with a much longer 36-inch f/10.0 aspheric lens. The B camera was a very complex device that used a single lens to obtain photography from one horizon to the other, thereby reducing weight by having two fewer lenses and shutter assemblies than the standard trimetrogon configuration. Because its lens was longer than those used in the A cameras, the B camera achieved even higher resolution—100 lines per millimeter.

The B camera used an 18- by 18-inch format, which was achieved by focusing the image onto two counterrotating but overlapping 9. 5-inch wide strips of film. Baker designed this camera so that one film supply was located forward, the other aft. Thus, as the film supplies unwound, they counterbalanced each other and did not disturb the aircraft's center of gravity.

The B camera had two modes of operation. In mode I, the camera used a single lens to make seven unique exposures from 73.5° on the far right and far left obliques to vertical photos beneath the aircraft, effectively covering from horizon to horizon. Mode II narrowed the lateral coverage to 21.5° on either side of vertical. This increased the available number of exposures and almost doubled the camera's

[18] "Basic Configuration and Camera Data," 24 January 1956, OSA Records (TS Codeword); *OSA History*, chap. 5, annex 44 (TS Codeword).



B camera

operating time. Three of the seven B-camera frames provided stereo coverage. The complex B cameras were engineered by Hycon's chief designer, William McFadden.[19]

James Baker's idea for the ultimate high-altitude camera was the C model that would have a 240-inch focal length. In December 1954, he made preliminary designs for folding the optical path using three mirrors, a prism, and an f/20.0 lens system. Before working out the details of this design, however, Baker flew to California in early January 1955 to consult with Kelly Johnson about the weight and space limitations of the U-2's payload compartment. Despite every effort to reduce the physical dimensions of the C camera, Baker needed an additional six inches of payload space to accommodate the bigger lens. When he broached this subject to Johnson, the latter replied, "Six more inches? I'd sell my grandmother for six more inches!"[20]

[19] Ibid.; Baker interview (S).

[20] Baker interview (S).

Realizing that the 240-inch lens was both too large and too heavy for the camera bay, Baker scaled the lens down to a 200-inch f/16. 0 system. This was still too big. Further reductions followed, resulting by July 1955 in a 120-inch f/10.9 lens that met both the weight and space limitations. Later in the year, Baker decided to make the mirrors for the system out of a new, lightweight foamed silica material developed by Pittsburgh-Corning Glass Company. This reduced the weight significantly, and he was able to scale up the lens to a 180-inch f/13.85 reflective system for a 13- by 13-inch format. In the past, the calculations for such a complex camera lens would have taken years to complete, but thanks to Baker's ray-tracing computer program, he was able to accomplish the task in just 16 days.

When a C camera built by Hycon was flight-tested on 31 January 1957, project engineers discovered that its 180-inch focal length, which was five times longer than that of the B camera, made the camera very sensitive to aircraft vibration and led to great difficulty in aiming the C camera from altitudes above 68,000 feet. The engineers, therefore, decided to shelve the camera. More than five years later, a redesigned C camera was employed during the Cuban Missile Crisis in October 1962, but the results were not very satisfactory.

The failure of the C camera design was not a serious setback to the high-altitude reconnaissance program, because the B camera proved highly successful. Once initial difficulties with the film-transport system were overcome, the B camera became the workhorse of high-altitude photography. An improved version known as the B-2 is still in use. Both of the earlier A-model cameras were phased out after September 1958.

During the period when he was designing lenses for the CIA's overhead reconnaissance program, James Baker was also working on classified lens designs for the Air Force and unclassified designs for the Smithsonian Institution. To protect the security of Baker's work for the Agency, Herbert Miller of the Development Projects Staff told Baker to work on lenses for the U-2 in the open and not make any effort to classify the documents connected with the project. Miller believed that by not calling attention to the effort through the use of special security measures, the project could be completed faster and still not be compromised. This "hiding in the open" strategy proved very successful.[21]

[21] Ibid.

In addition to the camera systems, the U-2 carried one other important item of optical equipment, a periscope. Designed by James Baker and built by Walter Baird of Baird Associates, the optical periscope helped pilots recognize targets beneath the aircraft and also proved to be a valuable navigational aid.[22]

PREPARATIONS FOR TESTING THE U-2

As work progressed in California on the airframe, in Connecticut on the engines, and in Boston on the camera system, the top officials of the Development Projects Staff flew to California and Nevada to search for a site where the aircraft could be tested safely and secretly. On 12 April 1955 Richard Bissell and Col. Osmund Ritland (the senior Air Force officer on the project staff) flew over Nevada with Kelly Johnson in a small Beechcraft plane piloted by Lockheed's chief test pilot, Tony LeVier. They spotted what appeared to be an airstrip by a salt flat known as Groom Lake, near the northeast corner of the Atomic Energy Commission's (AEC) Nevada Proving Ground. After debating about landing on the old airstrip, LeVier set the plane down on the lakebed, and all four walked over to examine the strip. The facility had been used during World War II as an aerial gunnery range for Army Air Corps pilots. From the air the strip appeared to be paved, but on closer inspection it turned out to have originally been fashioned from compacted earth that had turned into ankle-deep dust after more than a decade of disuse. If LeVier had attempted to land on the airstrip, the plane would probably have nosed over when the wheels sank into the loose soil, killing or injuring all of the key figures in the U-2 project.[23]

Bissell and his colleagues all agreed that Groom Lake would make an ideal site for testing the U-2 and training its pilots. Upon returning to Washington, Bissell discovered that Groom Lake was not part of the AEC proving ground. After consulting with Dulles, Bissell and Miller asked the Atomic Energy Commission to add the Groom Lake area to its real estate holdings in Nevada. AEC Chairman Adm. Lewis Strauss readily agreed, and President Eisenhower also approved the addition of this strip of wasteland, known by its map designation as Area 51, to the Nevada Test Site. The outlines of Area 51

[22] Information supplied by James Baker to Donald E. Welzenbach, 12 May 1986 (U).

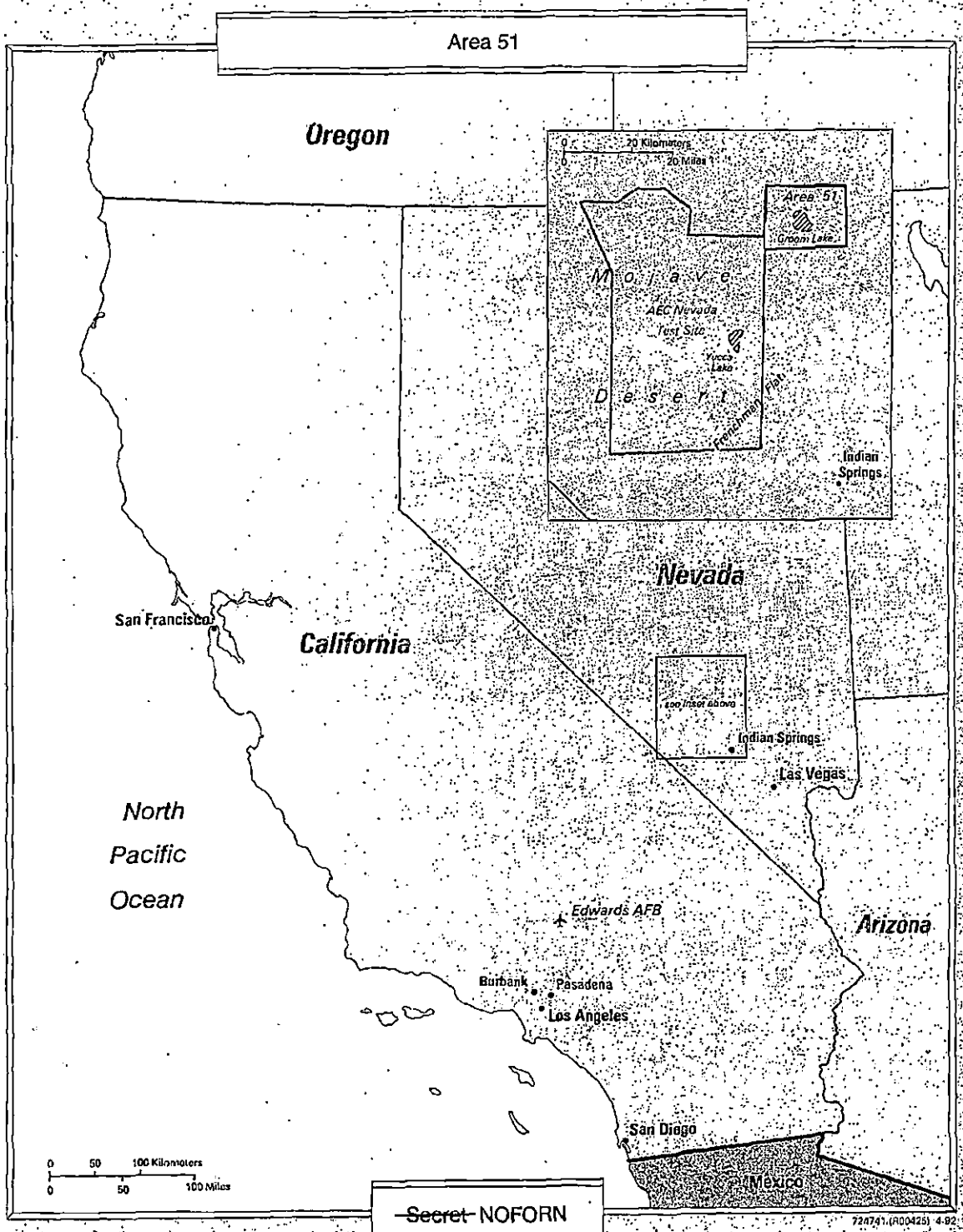
[23] *OSA History*, chap. 8, pp. 1-2 (TS Codeword); Miller, *Lockheed U-2*, pp. 19-20.

Area 51, the Ranch

are shown on current unclassified maps as a small rectangular area adjoining the northeast corner of the much larger Nevada Test Site. To make the new facility in the middle of nowhere sound more attractive to his workers, Kelly Johnson called it the Paradise Ranch, which was soon shortened to the Ranch.[24]

Although the dry lakebed could have served as a landing strip, project managers decided that a paved runway was needed so that testing could also take place during the times when rainwater runoff from nearby mountains filled the lake (at such times the base acquired yet another unofficial name, Watertown Strip). By July 1955 the base was ready, and Agency, Air Force, and Lockheed personnel began moving in.

[24] *OSA History*, chap. 8, pp. 2-6 (~~TS~~ Codeword); Johnson, "Log for Project X," 25-29 April 1955; Clarence L. "Kelly" Johnson with Maggie Smith, *Kelly: More Than My Share of It All* (Washington, DC: Smithsonian Institute Press, 1985), p. 123.


Area 51
Oregon
0 20 Kilometers
0 20 Miles
Area 51
Groom Lake
Mojave
AEC Nevada
Test Site
Yucca
Lake
Desert
Frenchman Flat
Indian
Springs
Nevada
San Francisco
California
see inset above
Indian Springs
Las Vegas
North
Pacific
Ocean
Edwards AFB
Arizona
Burbank
Pasadena
Los Angeles
San Diego
Mexico
0 50 100 Kilometers
0 50 100 Miles
~~Secret~~ NOFORN
72/741 (R00425) 4-92

SECURITY FOR THE U-2 PROJECT

On 29 April 1955, Richard Bissell signed an agreement with the Air Force and the Navy (which at that time was also interested in the U-2) in which the services agreed that the CIA "assumed primary responsibility for all security" for the overhead reconnaissance project (AQUATONE). From this time on, the CIA has been responsible for the security of overhead programs. This responsibility has placed a heavy burden on the Office of Security for establishing procedures to keep large numbers of contracts untraceable to the Central Intelligence Agency. The Office of Security has also had to determine which contractor employees require security clearances and has had to devise physical security measures for the various manufacturing facilities. Keeping the U-2 and subsequent overhead systems secret has been a time-consuming and costly undertaking.[25]

The most important aspect of the security program for the U-2 project was the creation of an entire new compartmented system for the product of U-2 missions. Access to the photographs taken by the U-2 would be strictly controlled, which often limited the ability of CIA analysts to use the products of U-2 missions.

The terminology used to describe U-2 aircraft and pilots also played a part in maintaining the security of the overhead reconnaissance program. To reduce the chances of a security breach, the Agency always referred to its high-altitude aircraft as "articles," with each aircraft having its own "article number." Similarly, the pilots were always called "drivers." In cable traffic the aircraft were known as KWEXTRA-00 (the two-digit number identified the precise aircraft; these numbers were not related to the three-digit article numbers assigned by the factory). The pilots were referred to as KWGLITTER-00 (the two-digit number identified the precise pilot). Thus, even if a message or document about overflight activities fell into unfriendly hands, the contents would simply refer to codewords or at worst to "articles" and "drivers," giving no indication of the nature of the program.[26]

Even the aircraft's onboard equipment required the involvement of CIA security planners. Thus, when Kelly Johnson ordered altimeters from the Kollman Instrument Company, he specified that the

[25] *OSA History*, chap. 7, pp. 4-6 (~~TS~~ Codeword).

[26] Information supplied by James Cunningham to Donald E. Welzenbach (~~S~~).

devices had to be calibrated to 80,000 feet. This immediately raised eyebrows at Kollman because its instruments only went to 45,000 feet. Agency security personnel quickly briefed several Kollman officials and produced a cover story that the altimeters were to be used on experimental rocket planes.[27]

THE CIA – AIR FORCE PARTNERSHIP

At the initial interagency meetings to establish the U-2 program in December 1954, the participants did not work out a clear delineation of responsibilities between the CIA and the Air Force. They agreed only that the Air Force would supply the engines and the Agency would pay for the airframes and cameras. With a myriad of details still unsettled, CIA and Air Force representatives began to work on an interagency agreement that would assign specific responsibilities for the program. These negotiations proved difficult. Discussions on this subject between DCI Allen Dulles and Air Force Chief of Staff Nathan Twining began in March 1955. Twining wanted SAC, headed by Gen. Curtis E. LeMay, to run the project once the planes and pilots were ready to fly, but Dulles opposed such an arrangement. The CIA-USAF talks dragged on for several months, with Twining determined that SAC should have full control once the aircraft was deployed. Eventually President Eisenhower settled the dispute. "I want this whole thing to be a civilian operation," the President wrote. "If uniformed personnel of the armed services of the United States fly over Russia, it is an act of war—legally—and I don't want any part of it."[28]

With the issue of control over the program settled, the two agencies soon worked out the remaining details. On 3 August 1955, Dulles and Twining met at SAC headquarters in Omaha to sign the basic agreement, titled "Organization and Delineation of Responsibilities—Project OILSTONE" (OILSTONE was the Air Force codename for the project). This pact gave the Air Force responsibility for pilot selection and training, weather information, mission plotting, and operational support. The Agency was responsible for cameras, security, contracting, film processing, and arrangements for foreign bases, and it also had a voice in the selection of pilots. All aeronautical aspects of the

[27] Ibid.

[28] *OSA History*, chap. 3, pp. 8-15 (~~TS~~ Codeword); Beschloss, *Mayday*, pp. 105-107.

project—the construction and testing of the aircraft—remained the exclusive province of Lockheed.[29]

As a result of this agreement, CIA remained in control of the program, but the Air Force played a very important role as well. As Richard Bissell later remarked about the U-2 project, "The Air Force wasn't just in on this as a supporting element, and to a major degree it wasn't in on it just supplying about half the government personnel; but the Air Force held, if you want to be precise, 49 percent of the common stock."[30]

Osmund J. Ritland

One of the first Air Force officers assigned to Project OILSTONE was Col. Osmund J. Ritland. He began coordinating Air Force activities in the U-2 program with Richard Bissell in December 1954. On 27 June 1955, Ritland became Bissell's deputy, although Air Force Chief of Staff Twining did not officially approve this assignment until 4 August, the day after the signing of the CIA–Air Force agreement. In March 1956, Colonel Ritland returned to the Air Force and was followed as deputy project director by Col. Jack A. Gibbs.

Another Air Force officer, Lt. Col. Leo P. Geary, joined the program in June 1955 and remained with it until August 1966, longer than any of the other project managers. Using the Air Force Inspector General's office as cover with the title of Project Officer, AFCIG-5, Geary served as the focal point for all Defense Department support to the U-2 and OXCART programs. His 11 years with the overhead reconnaissance projects provided a high degree of Air Force continuity.[31]

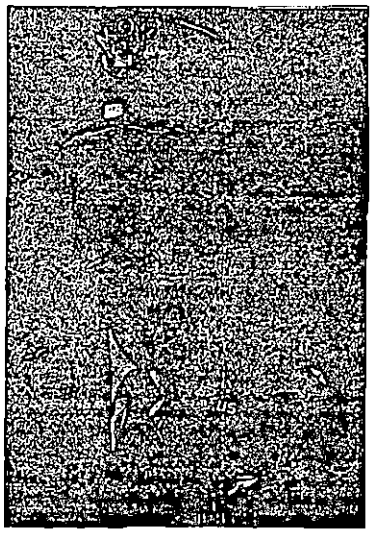
TECHNICAL CHALLENGES TO HIGH-ALTITUDE FLIGHT

To get the U-2 aircraft ready to fly, Lockheed engineers had to solve problems never before encountered. Among these problems was the need for a fuel that would not boil off and evaporate at the very high altitudes for which the aircraft was designed. Gen. James H. Doolittle

[29] *OSA History*, chap. 3, p. 15 and annex 14 (~~CPS~~ Codeword).

[30] Speech given by Richard Bissell at CIA Headquarters, 12 October 1965 (~~CPS~~ Codeword).

[31] Brig. Gen. Leo A. Geary (USAF-Ret.), interview by Donald E. Welzenbach, tape recording, 3 April 1986 (S); *OSA History*, chap. 3, p. 3 (~~CPS~~ Codeword).



Leo P. Geary

(USAF, Ret.), a vice president of the Shell Oil Company who had long been involved in overhead reconnaissance (most recently as a member of the Technological Capabilities Panel), arranged for Shell to develop a special low-volatility, low-vapor-pressure kerosene fuel for the craft. The result was a dense mixture, known as LF-1A, JP-TS (thermally stable), or JP-7, with a boiling point of 300°F at sea level. Manufacturing this special fuel required petroleum byproducts that Shell normally used to make its "Flit" fly and bug spray. In order to produce several hundred thousand gallons of LF-1A for the U-2 project in the spring and summer of 1955, Shell had to limit the production of Flit, causing a nationwide shortage. Because of the new fuel's density, it required special tanks and modifications to the aircraft's fuel-control and ignition systems.[32]

Even more important than the problem of boiling fuel was the problem of boiling blood, namely the pilot's. At altitudes above 65,000 feet, fluids in the human body will vaporize unless the body can be kept under pressure. Furthermore, the reduced atmospheric pressure placed considerable stress on the pilot's cardiovascular system and did not provide adequate oxygenation of the blood. Keeping the pilot alive at the extreme altitudes required for overflights therefore called for a totally different approach to environmental equipment; it required a system that could maintain pressure over much of the pilot's body. The technology that enabled U-2 pilots to operate for extended periods in reduced atmospheric pressure would later play a major role in the manned space program.

Advising the Agency on high-altitude survival were two highly experienced Air Force doctors, Col. Donald D. Flickinger and Col. W. Randolph Lovelace, II. Dr. Lovelace had begun his research on high-altitude flight before World War II and was a coinventor of the standard Air Force oxygen mask. In the early 1950s, he and Flickinger made daring parachute jumps from B-47 bombers to test pilot-survival gear under extreme conditions. Flickinger served as the medical adviser to Project AQUATONE for almost a decade.[33]

Flickinger and Lovelace suggested that the Agency ask the David Clark Company of Worcester, Massachusetts, manufacturer of environmental suits for Air Force pilots, to submit designs for more

[32] Land interview (TS Codeword); Bissell interview (S); James A. Cunningham, Jr., interview by Donald E. Welzenbach, Washington, DC, tape recording, 4 October 1983 (~~TS~~ Codeword).

[33] OSA History, chap. 10, pp. 29-34 (~~TS~~ Codeword).

MC-2 partial-pressure suit (seen on pilot Francis Gary Powers)

advanced gear for the U-2 pilots. David Clark expert Joseph Ruseckas then developed a complex life-support system, which was the first partially pressurized "spacesuit" for keeping humans alive for lengthy periods at ultrahigh altitudes. The effort to provide a safe environment for pilots at high altitudes also involved the Firewel Company of Buffalo, New York, which pressurized the U-2 cockpit to create an interior environment equivalent to the air pressure at an altitude of 28,000 feet. The system was designed so that, if the interior cockpit pressure fell below the 28,000-feet level, the pilot's suit would automatically inflate. In either case, he could obtain oxygen only through his helmet.[34]

[34] Ibid., chap. 5, p. 19 (~~TS~~ Codeword).

The early models of these MC-2 and MC-3 partial-pressure suits were very uncomfortable for the pilots. To prevent loss of pressure, the heavy coverall had to fit tightly at the wrists and ankles (in the early models of these suits, the feet were not included in the pressurization scheme). The pilot had to wear gloves and a heavy helmet that tended to chafe his neck and shoulders and was prone to fogging. Problems with the pilot life-support system were believed to have been the cause of several early crashes of the U-2.

Having gotten a pilot into this bulky suit and shoehorned him into his seat in the cockpit, the next problem was how to get him out in an emergency. The U-2 cockpit was very small, and the early models did not have an ejection seat. Even after an ejection seat was installed, pilots were reluctant to use it because they were afraid they would lose their legs below the knees when they were blown out of the cockpit. To save weight, the first pilot's seat was extremely simple with no height adjustment mechanism. Designed for pilots of above-average height, the seat could be adjusted for shorter pilots by inserting wooden blocks beneath the seat to raise it. In later versions of the aircraft, Kelly Johnson added a fully adjustable seat.[35]

The Air Force undertook bailout experiments at high altitudes from balloons in the autumn of 1955 to determine if the suit designed for the U-2 pilot would also protect him during his parachute descent once he was separated from the life-support mechanisms inside the aircraft. To avoid getting the "bends" during such descents or during the long flights, pilots had to don their pressure suits and begin breathing oxygen at least 90 minutes before takeoff so that their bodies would have time to dissipate nitrogen. This procedure was known as prebreathing. Once the pilots were in their suits, eating and drinking became a major problem, as did urination. The first model of the pressure suit, used by Lockheed test pilots, made no provision for urination. A subsequent model required the pilot to be catheterized before donning his flying suit. This method of permitting urination during flight proved very uncomfortable and, by the autumn of 1955, was replaced with an external bladder arrangement that made the catheter unnecessary. To reduce elimination, pilots ate a low-bulk, high-protein diet on the day before and the morning of each mission.

[35] Lecture by Maj. Gen. Patrick J. Halloran (former Air Force U-2 pilot) at the National Air & Space Museum, 24 April 1986 (U).

Pilot undergoing prebreathing

To prevent pilots from becoming dessicated during the long missions—a condition aggravated by their having to breathe pure oxygen—provision was made for them to drink sweetened water. This was accomplished by providing a small self-sealing hole in the face mask through which the pilot could push a strawlike tube attached to the water supply. Project personnel also pioneered in the development of ready-to-eat foods in squeezable containers. These were primarily bacon- or cheese-flavored mixtures that the pilot could squeeze into his mouth using the self-sealing hole in the face mask. Despite all these precautions, U-2 pilots normally lost 3 to 6 pounds of body weight during an eight-hour mission.[36]

Food and water were not the only items provided to pilots on overflight missions; they also received a suicide pill. During the early 1950s, tales of Soviet secret police torture of captured foreign agents

[36] Information supplied by James Cunningham and former U-2 pilots Carmine Vito, Hervey Stockman, Jacob Kratt, and Glendon Dunaway to Donald E. Welzenbach, May 1986.

led Bissell and Cunningham to approach Dr. Alex Batlin of Technical Services Division in the Directorate of Plans[37] for ideas to help "captured" U-2 pilots avoid such suffering. Batlin suggested the method used by Nazi war criminal Hermann Goering, a thin glass ampule containing liquid potassium cyanide. He said a pilot had only to put the ampule in his mouth and bite down on the glass; death would follow in 10 to 15 seconds. Project AQUATONE ordered six of the poison ampules, called L-pills, and offered one to each pilot just before a mission. It was up to each pilot to decide if he wanted to take an L-pill with him. Some did; most did not.[38]

DELIVERY OF THE FIRST U-2

On 25 July, less than eight months after the go-ahead call from Trevor Gardner, Kelly Johnson was ready to deliver the first aircraft, known as article 341, to the "Paradise Ranch" site. With its long, slender wings and tail assembly removed, the aircraft was wrapped in tarpaulins, loaded aboard a C-124, and flown to Groom Lake, where Lockheed mechanics spent the next six days readying the craft for its maiden flight.

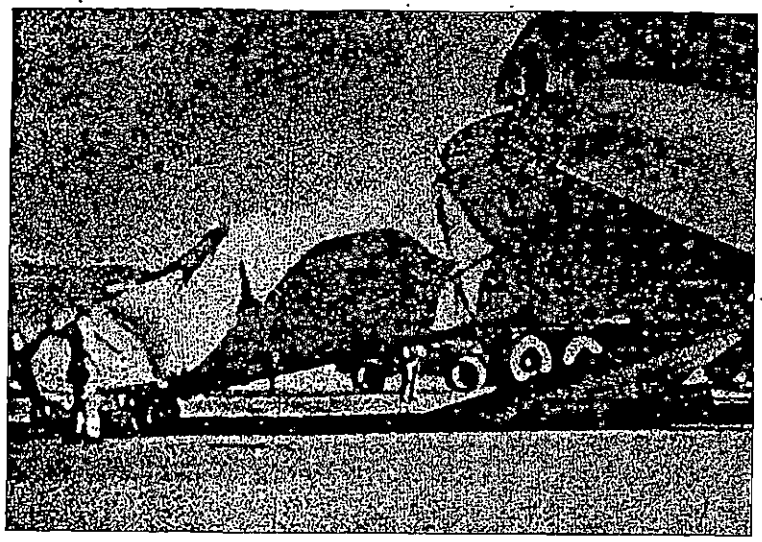
Before "Kelly's Angel" could actually take to the air, however, it needed an Air Force designator. Col. Allman T. Culbertson from the Air Force's Office of the Director of Research and Development pointed this out to Lieutenant Colonel Geary in July 1955, and the two officers then looked through the aircraft designator handbook to see what the options were. They decided that they could not call the project aircraft a bomber, fighter, or transport plane, and they did not want anyone to know that the new plane was for reconnaissance, so Geary and Culbertson decided that it should come under the utility aircraft category. At the time, there were only two utility aircraft on the books, a U-1 and a U-3. Geary told Culbertson that the Lockheed CL-282 was going to be known officially as the U-2.[39]

---

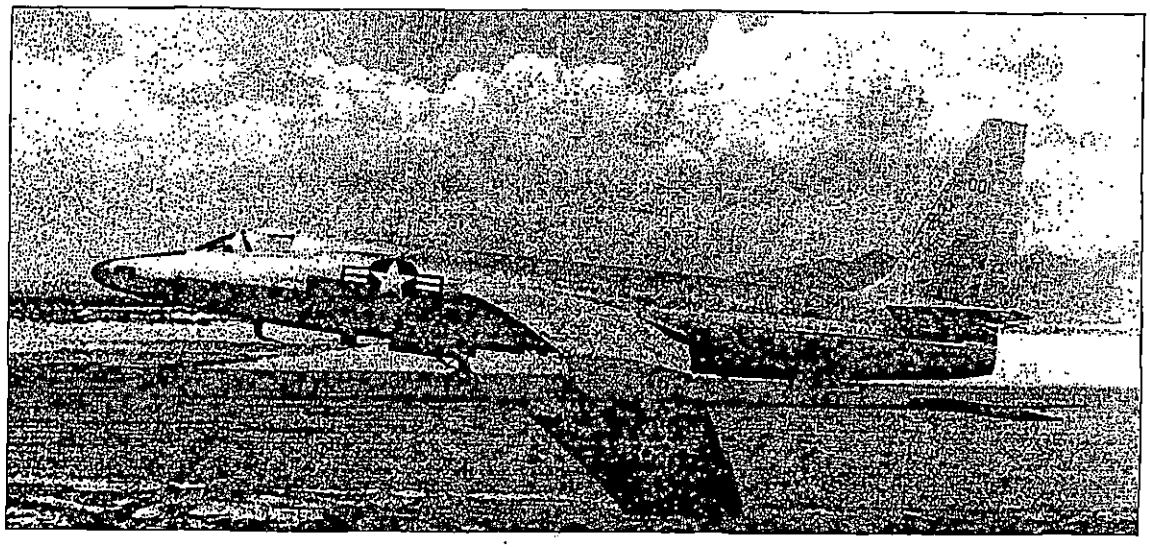
[37] At the time this Directorate was known as the Deputy Directorate/Plans, with the slash interpreted to mean either "for" or "of." Terminology for the major subdivisions of the CIA and their directors has varied over the past four decades. For the sake of consistency, all titles of Directorates and Deputy Directors have been placed in the current Agency format: the organization is known as the "Directorate of X" and the head is known as the "Deputy Director for X."

[38] Information supplied by James Cunningham to Donald E. Welzenbach; Sayre Stevens, Memorandum for the Record, "Discussion with Dr. Alex Batlin Re Project MKNAOMI," July 1975 (~~S~~).

[39] Geary interview (~~S~~).



Arrival of U-2 prototype at Area 51 (left); Article 341, the U-2 prototype (below)



Johnson had designed the U-2 to use the Pratt & Whitney (P&W) J57/P-31 engine, which developed 13,000 pounds of thrust and weighed 3,820 pounds, giving it a power-to-weight ratio of 3.4:1. When the U-2 first took to the air, however, these engines were not available because the entire production was needed to power specially configured Canberra RB-57Ds for the Air Force. The first U-2s therefore used P&W J57/P-37 engines, which were 276 pounds heavier and delivered only 10,200 pounds of thrust at sea level; the resulting

power-to-weight ratio of 2.7:1 was almost 20 percent less efficient than the preferred P-31 version.[40]

To conduct lengthy missions over hostile territory, the U-2 needed to carry a large amount of fuel. Kelly Johnson used a "wet-wing" design for the U-2, which meant that fuel was not stored in separate fuel tanks but rather in the wing itself. Each wing was divided into two leak-proof compartments, and fuel was pumped into all the cavities within these areas; only the outer 6 feet of the wings were not used for fuel storage. The U-2 also had a 100-gallon reserve tank in its nose. Later, in 1957, Johnson increased the fuel capacity of the U-2 by adding 100-gallon "slipper" tanks under each wing, projecting slightly ahead of the leading edge.

One of the most important considerations in the U-2's fuel system was the need to maintain aircraft trim as the fuel was consumed. The aircraft therefore contained a complex system of feed lines and valves draining to a central sump, which made it impossible to provide the pilot with an empty/full type of fuel gauge. None of the first 50 U-2s had normal fuel gauges. Instead there were mechanical fuel totalizer/counters. Before the start of a mission, the ground crew set the counters to indicate the total amount of fuel in the wings, and then a flow meter subtracted the gallons of fuel actually consumed during the flight. The pilot kept a log of the fuel consumption shown by the counters and compared it with estimates made by mission planners for each leg of the flight. As a double check, U-2 pilots also kept track of their fuel consumption by monitoring airspeed and time in the air. Most pilots became quite expert at this. Several who did not came up short of their home base during the 20 years these planes were flown.[41]

INITIAL TESTING OF THE U-2

Preliminary taxi trials began on 27 July 1955, when the first run down the newly completed runway took the plane to 50 knots. Lockheed's chief test pilot, Tony LeVier, was at the controls. A second taxi trial

[40] *OSA History*, chap. 8, p. 13 (~~TS~~ Codeword).

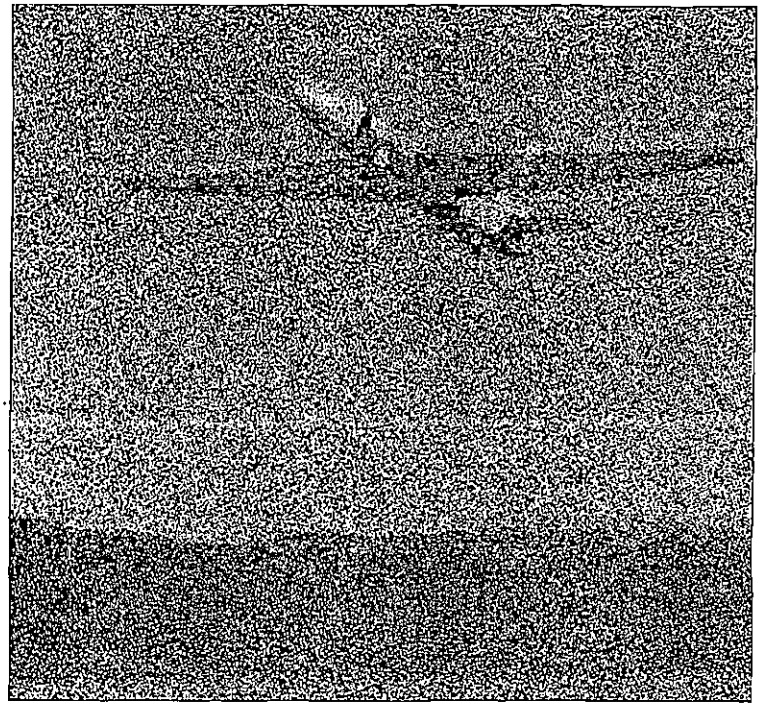
[41] Information supplied by Norman Nelson, former director of Lockheed's Skunk Works, to Donald E. Welzenbach, 14 March 1986 (U); Miller, *Lockheed U-2*, pp. 77, 96.

followed on 1 August. LeVier accelerated to 70 knots and began to try the ailerons. "It was at this point that I became aware of being airborne," LeVier noted afterward, "which left me with utter amazement, as I had no intentions whatsoever of flying. I immediately started back toward the ground, but had difficulty determining my height because the lakebed had no markings to judge distance or height. I made contact with the ground in a left bank of approximately 10 degrees." The U-2 bounced back into the air, but LeVier was able to bring it back down for a second landing. He then applied the brakes with little effect, and the aircraft rolled for a long distance before coming to a stop.[42]

Bissell, Cunningham, and Johnson saw the aircraft fall and bounce. Leaping into a jeep, they roared off toward the plane. They signaled to LeVier to climb out and then used fire extinguishers to put out a fire in the brakes. At a debriefing session that followed, LeVier complained about the poor performance of the brakes and the absence of markings on the runway. Damage to the prototype U-2 was very minor: blown tires, a leaking oleostrut on the undercarriage, and damaged brakes. This unplanned flight was but a foretaste of the airworthiness of the U-2. New pilots all had difficulty in getting the U-2's wheels on the ground because at low speeds it would remain in ground effect and glide effortlessly above the runway for great distances.

Taxi trials continued for one more day and were followed by the first planned flight on 4 August 1955. LeVier was again at the controls and had been instructed by Kelly Johnson to land the U-2 by making initial contact with the main or forward landing gear and letting the plane settle back on the rear wheel. LeVier had disagreed with this approach, believing that the U-2 would bounce if he tried to touch down on the forward gear first. After flying the aircraft up to 8,000 feet, LeVier leveled off and began cycling the landing gear up and down; then he tested the flaps and the plane's stability and control systems. Finally, LeVier made his first landing approach. As the U-2 settled down, the forward landing gear touched the runway and the plane skipped and bounced into the air. LeVier made a second attempt to land front wheels first, and again the plane bounded into the air.

[42] Transcripts of the test pilots' and observers' comments on the initial U-2 test flights have been published in "Secret First Flight of Article 001," *Spyplanes* vol. 2, 1988, pp. 64-71, 82-85.



First flight of the U-2, 4 August 1955

With Kelly Johnson watching from a chase plane and giving a constant stream of instructions, LeVier made three more unsuccessful landing attempts. With the light fading and a thunderstorm fast approaching from the mountains to the west, LeVier made one last approach using the method he had advocated: letting the aircraft touch on its rear wheel first. This time the U-2 made a near-perfect landing, which came just in the nick of time. Ten minutes later, the thunderstorm began dumping an unheard-of 2 inches of rain, flooding the dry lakebed and making the airstrip unusable.[43]

Now that the first problems in flying and landing the U-2 had been worked out, Kelly Johnson scheduled the "official" first flight for 8 August 1955. This time outsiders were present, including Richard Bissell, Col. Osmond Ritland, Richard Horner, and Garrison Norton. The U-2 flew to 32,000 feet and performed very well. Kelly Johnson had met his eight-month deadline.[44]

[43] Ibid., pp. 21-22; Johnson, "Log for Project X," 4 August 1955.

[44] Johnson, "Log for Project X," 8 August 1955.

LeVier made an additional 19 flights in article 341 before moving on to other Lockheed flight test programs in early September. This first phase of U-2 testing explored the craft's stall envelope, took the aircraft to its maximum stress limit (2.5 g's), and explored its speed potential. LeVier soon flew the aircraft at its maximum speed of Mach 0.85. Flight tests continued, with the U-2 ascending to altitudes never before attained in sustained flight. On 16 August LeVier took the aircraft up to 52,000 feet. In preparation for this flight, the 42-year-old test pilot completed the Air Force partial-pressure suit training program, becoming the oldest pilot to do so. Testing at even higher altitudes continued, and on 8 September the U-2 reached its initial design altitude of 65,600 feet.[45]

On 22 September 1955, the U-2 experienced its first flameout at 64,000 feet—more than 12 miles up. After a brief restart, the J57/P-37 engine again flamed out at 60,000 feet, and the aircraft descended to 35,000 feet before the engine could be relit. Engineers from Pratt & Whitney immediately set to work on this problem. The P-37 model engine had significantly poorer combustion characteristics than the preferred but unavailable P-31 version and therefore tended to flame out at high altitudes. Combustion problems usually became apparent as the U-2 began the final part of its climb from 57,000 to 65,000 feet, causing pilots to refer to this area as the "badlands" or the "chimney." Flameouts bedeviled the U-2 project until sufficient numbers of the more powerful P-31 engines became available in the spring of 1956.[46]

Meanwhile, with the airworthiness of the U-2 airframe proven, Lockheed set up a production line in the Skunk Works, but delivery of even the second-choice J57/P-37 engines became a major problem. Pratt & Whitney's full production capacity for these engines for the next year was contracted to the Air Force for use in F-100 fighters and KC-135 tankers. Colonel Geary, with the help of a colleague in the Air Force Materiel Command, managed to arrange the diversion of a number of these engines from a shipment destined for Boeing's KC-135 production line, making it possible to continue building the U-2s.[47]

---

[45] *OSA Chronology*, p. 7 (TS Codeword); Miller, *Lockheed U-2*, p. 22.

[46] *OSA History*, chap. 8, pp. 12-14 (~~TS~~ Codeword).

[47] Geary interview (~~S~~).

As the deliveries of U-2 airframes to the testing site increased, a major logistic problem arose: how to transfer Lockheed employees from Burbank to Area 51 without arousing a great deal of curiosity. The project staff decided that the simplest approach would be to fly the essential personnel to the site on Monday morning and return them to Burbank on Friday evening. Frequent flights were also necessary to bring in supplies and visitors from contractors and headquarters. Therefore, a regularly scheduled Military Air Transport Service (MATS) flight using a USAF C-54 aircraft began on 3 October 1955. James Cunningham promptly dubbed this activity "Bissell's Narrow-Gauge Airline." Less than seven weeks after it started, a MATS aircraft bound for Area 51 crashed on 17 November, killing all 14 persons aboard the plane, including the Project Security Officer, CIA's William H. Marr, four members of his staff, and personnel from Lockheed and Hycon. This crash represented the greatest single loss of life in the entire U-2 program.[48]

U-2s, UFOs, AND OPERATION BLUE BOOK

High-altitude testing of the U-2 soon led to an unexpected side effect—a tremendous increase in reports of unidentified flying objects (UFOs). In the mid-1950s, most commercial airliners flew at altitudes between 10,000 and 20,000 feet and military aircraft like the B-47s and B-57s operated at altitudes below 40,000 feet. Consequently, once U-2s started flying at altitudes above 60,000 feet, air-traffic controllers began receiving increasing numbers of UFO reports.

Such reports were most prevalent in the early evening hours from pilots of airliners flying from east to west. When the sun dropped below the horizon of an airliner flying at 20,000 feet, the plane was in darkness. But, if a U-2 was airborne in the vicinity of the airliner at the same time, its horizon from an altitude of 60,000 feet was considerably more distant, and, being so high in the sky, its silver wings would catch and reflect the rays of the sun and appear to the airliner pilot, 40,000 feet below, to be fiery objects. Even during daylight hours, the silver bodies of the high-flying U-2s could catch the sun and cause reflections or glints that could be seen at lower altitudes and even on the ground. At this time, no one believed manned flight was possible above 60,000 feet, so no one expected to see an object so high in the sky.

[48] *OSA History*, chap. 7, pp. 17-19 (~~TS~~ Codeword).

Not only did the airline pilots report their sightings to air-traffic controllers, but they and ground-based observers also wrote letters to the Air Force unit at Wright Air Development Command in Dayton charged with investigating such phenomena. This, in turn, led to the Air Force's Operation BLUE BOOK. Based at Wright-Patterson, the operation collected all reports of UFO sightings. Air Force investigators then attempted to explain such sightings by linking them to natural phenomena. BLUE BOOK investigators regularly called on the Agency's Project Staff in Washington to check reported UFO sightings against U-2 flight logs. This enabled the investigators to eliminate the majority of the UFO reports, although they could not reveal to the letter writers the true cause of the UFO sightings. U-2 and later OXCART flights accounted for more than one-half of all UFO reports during the late 1950s.[49]

HIRING U-2 PILOTS

In authorizing the U-2 project, President Eisenhower told DCI Dulles that he wanted the pilots of these planes to be non-US citizens. It was his belief that, should a U-2 come down in hostile territory, it would be much easier for the United States to deny any responsibility for the activity if the pilot was not an American.

25X1 and 6, E.O.13526

The initial effort to find U-2 pilots was assigned to the Directorate of Plans Air/Maritime Division (AMD). The DDP had excellent contacts [redacted] AMD operatives, [redacted] asked that discreet inquiries be made to see if any US-trained fliers were interested in a high-paying covert project. [redacted] brought to the United States for training. Meanwhile, AMD hired an [redacted] flier residing in England, and he also came to the United States for training.

In theory the use of foreign pilots seemed quite logical; in practice it did not work out. [redacted] and could only fly light aircraft. Language was also a barrier for the [redacted] although several were good fliers. Because Lieutenant Colonel Geary had taken a class of [redacted] through flying school at Randolph AFB in 1950, he got the job of training the [redacted]

25X1 and 6, E.O.13526

[49] Information supplied by James Cunningham to Donald E. Welzenbach (U).

25X1 and 6, E.O.13526

recruits in mid-1955. Geary arranged for an Air Force officer of [redacted] to stay with the group during a preliminary training program at Luke Air Force Base. The plan to use foreign pilots soon ran into trouble when only [redacted] pilots passed the school and reported to Area 51. They made only a few flights in the U-2, and by the autumn of 1955 they were out of the program.[50]

25X1 and 6, E.O.13526

Even before the elimination of the [redacted] it was clear that there would not be enough trained foreign pilots available in time for deployment. Bissell therefore had to start the search for U-2 pilots all over again. Lt. Gen. Emmett (Rosy) O'Donnell, the Air Force's Deputy Chief of Staff for Personnel, authorized the use of Air Force pilots and provided considerable assistance in the search for pilots who met the high standards established by the Agency and the Air Force. The search included only SAC fighter pilots who held reserve commissions. The use of regular Air Force pilots was not considered because of the complexities involved in having them resign from the Air Force, a procedure that was necessary in order to hire them as civilians for the AQUATONE project.

SAC pilots interested in the U-2 project had to be willing to resign from the Air Force and assume civilian status—a process known as sheep-dipping—in order to conduct the overflights. Although Air Force pilots were attracted by the challenge of flying U-2s over hostile territory, they were reluctant to leave the service and give up their seniority. To overcome pilots' reluctance, the Agency offered handsome salaries, and the Air Force promised each pilot that, upon satisfactory conclusion of his employment with the Agency, he could return to his unit. In the meantime, he would be considered for promotion along with his contemporaries who had continued their Air Force careers.[51]

The selection process for Agency U-2 pilots was very rigorous. Because of the strain involved in flying at extreme altitudes for long periods of time, painstaking efforts were made to exclude all pilots who might be nervous or unstable in any way. The physical and psychological screening of potential U-2 pilots was conducted by the Lovelace Foundation for Medical Education and Research in Albuquerque, New Mexico, under a contract signed with the CIA on

[50] *OSA History*, chap. 10, pp. 1-10 (~~TS~~ Codeword); Geary interview ~~(S)~~.

[51] *OSA History*, chap. 10, pp. 5-6 (~~TS~~ Codeword); Geary interview (S); Francis Gary Powers with Curt Gentry, *Operation Overflight* (New York: Holt, Rinehart, and Wilson, 1970), pp. 25-27.

28 November 1955. The CIA's insistence on more stringent physical and mental examinations than those used by the Air Force to select pilots for its U-2 fleet resulted in a higher rejection rate of candidates. The Agency's selection criteria remained high throughout its manned overflight program and resulted in a much lower accident rate for CIA U-2 pilots than for their counterparts in the Air Force program.[52]

PILOT TRAINING

Even before the recruiting effort got under way, the Air Force and CIA began to develop a pilot training program. Under the terms of the OILSTONE agreement between the Agency and the Air Force, responsibility for pilot training lay with SAC. This essential activity was carried out under the supervision of Col. William F. Yancey, who was assigned to March AFB and flew to nearby Area 51 each day. Colonel Yancey was in charge of six SAC pilots who were to be trained by Lockheed test pilots to fly the U-2. Once they became qualified, these SAC pilots would become the trainers for the "sheep-dipped" former Reserve SAC pilots, who would fly U-2 missions for the CIA.

The original U-2 test pilot, Tony LeVier, trained several other Lockheed test pilots in the difficult art of flying the U-2. Eventually there were enough trained Lockheed pilots available to test the aircraft coming off the assembly line and also train the SAC pilots. Training was difficult because there was no two-seat model of the U-2. All instruction had to be given on the ground before takeoff and then over the radio once the craft was airborne. Almost 15 years elapsed before a two-seat U-2 was available for training new pilots. Despite the difficulties involved in training U-2 pilots, Colonel Yancey had a cadre of six qualified Air Force U-2 pilots by September 1955. These six were now ready to train the Agency's pilots.[53]

Training pilots was not easy because the U-2 was a mixture of glider and jet. Although those chosen for the overflight program were all qualified fighter pilots, they now had to learn to fly the delicate U-2. Its large wings had tremendous lift but were also very fragile

[52] *OSA History*, chap. 10, pp. 5-6; chap. 5, p. 18 (~~TS~~ Codeword).

[53] *OSA History*, chap. 11, pp. 1-7 (~~TS~~ Codeword).

and could not survive the stresses of loops and barrel rolls. Moreover, the original U-2s were placarded, which meant that they could not be flown at sea level faster than 190 knots in smooth air or 150 knots in rough air. At operational altitude, where the air was much less dense, they could not exceed Mach 0.8 (394 knots). Speeds in excess of these limits could cause the wings or tail section to fall off.

Airspeed was a very critical factor for the U-2. At maximum altitude only 6 knots separated the speeds at which low-speed stall and high-speed buffet occurred. Pilots called this narrow range of acceptable airspeeds at maximum altitude the "coffin corner" because at this point the U-2 was always on the brink of falling out of the sky. If the aircraft slowed beyond the low-speed stall limit, it would lose lift and begin to fall, causing stresses that would tear the wings and tail off. A little too much speed would lead to buffeting, which would also cause the loss of the wings or tail. Flying conditions such as these required a U-2 pilot's full attention when he was not using the autopilot. Airspeed was such a critical factor that Kelly Johnson added a vernier adjustment to the throttle to allow the pilot to make minute alterations to the fuel supply.[54]

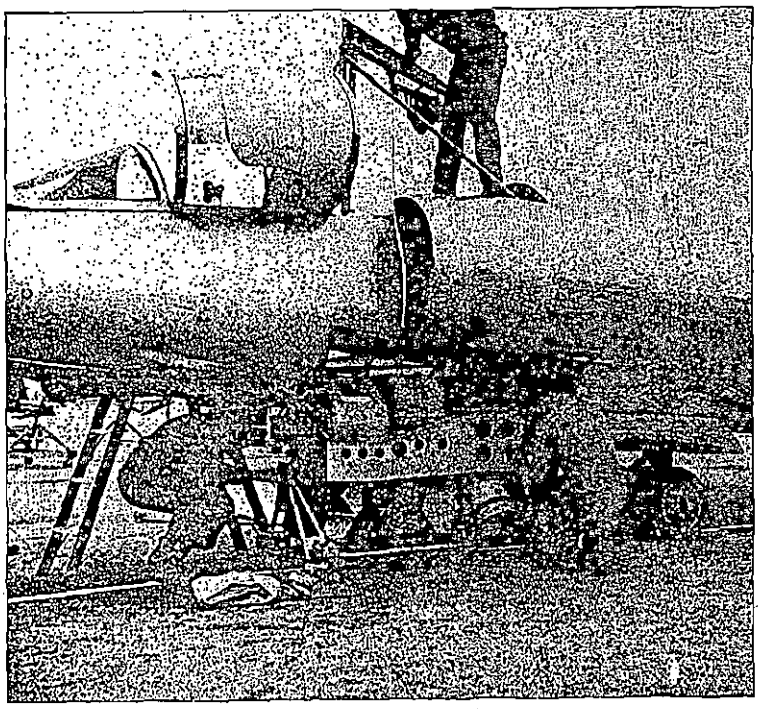
Among the unique devices developed for the U-2 was a small sextant for making celestial "fixes" during the long overflights. Because cloud cover often prevented U-2 pilots from locating navigational points on the earth through the periscope, the sextant turned out to be the pilots' principal navigational instrument during the first three years of deployment. When clouds were not a factor, however, the periscope proved highly accurate for navigation. During the final tests before the aircraft became operational, U-2 pilots found they could navigate by dead reckoning with an error of less than 1 nautical mile over a 1,000-nm course.[55]

FINAL TESTS OF THE U-2

Flight-testing of the U-2 continued throughout the fall and winter of 1955-56 in order to test all the various systems. By mid-January 1956, SAC officials were so impressed that they also wanted to purchase a fleet of these planes. On 30 January, DCI Dulles agreed to

[54] Cunningham interview (~~TS~~ Codeword); John Parangosky, interview by Donald E. Welzenbach, tape recording, 6 March 1986 (~~S~~); information supplied by James Cherbonneaux to Donald E. Welzenbach (S).

[55] Cunningham interview (~~TS~~ Codeword).

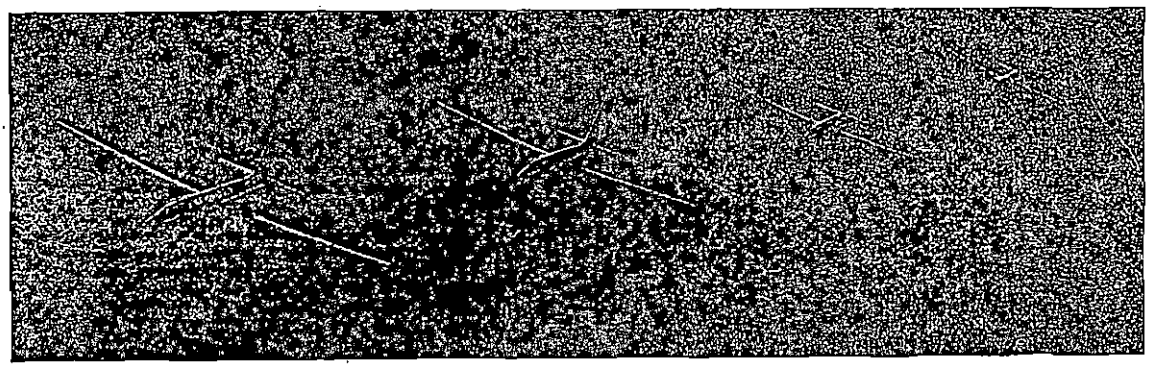


A-2 camera being installed in U-2

have CIA act as executive agent for this transaction, which the Air Force called Project DRAGON LADY. To maintain secrecy, the Air Force transferred funds to the CIA, which then placed an order with Lockheed for 29 U-2s in configurations to be determined by the Air Force. The Air Force later bought two more U-2s, for a total of 31. The aircraft purchased for the Air Force were known as the Follow-On Group, which was soon shortened to FOG.[56]

Once enough pilots had been trained, Project AQUATONE managers concentrated on checking out the complete U-2 system: planes, pilots, navigation systems, life-support systems, and cameras. From 10 through 14 April 1956, U-2s equipped with A-2 cameras took off from Area 51 and made eight overflights of the United States in order to test the various flight and camera systems as part of the standard Air Force Operational Readiness Inspection. Colonel Yancey and his detachment served as observers during this weeklong exercise.

[56] *OSA History*, chap. 5, pp. 25-26 (~~TS~~ Codeword).



U-2 detachment in formation over Nevada

Colonel Yancey's group carefully examined all aspects of the U-2 unit from flight crews to camera technicians and mission programmers. When the exercise was over, Yancey reported that the detachment was ready for deployment. He then briefed a high-level Pentagon panel that included the Secretary of the Air Force and the Chief of Air Staff. These officials concurred with Yancey's determination that the U-2 was ready for deployment.[57]

During these final tests in the spring of 1956, the U-2 once again demonstrated its unique airworthiness. On 14 April 1956, James Cunningham was sitting in his office in Washington when he received a call from Area 51 informing him that a westward-bound U-2 had experienced a flameout over the Mississippi River at the western border of Tennessee. After restarting his engine, the pilot reported a second flameout and engine vibrations so violent that he was unable to get the power plant to start again. Early in the program Bissell and Ritland had foreseen such an emergency and, with the cooperation of the Air Force, had arranged for sealed orders to be delivered to every airbase in the continental United States giving instructions about what to do if a U-2 needed to make an emergency landing.

Cunningham had the project office ask the pilot how far he could glide so they could determine which SAC base should be alerted. The pilot, who by this time was over Arkansas, radioed back that, given the prevailing winds and the U-2's 21:1 glide ratio, he thought he could reach Albuquerque, New Mexico. Within minutes Cunningham was on the phone to Colonel Geary in the Pentagon, who then had the Air Force's Assistant Director of Operations, Brig. Gen. Ralph E.

[57] Bissell interview (~~S~~); *OSA History*, chap. 11, pp. 15-16 (~~TS~~ Codeword).

Koon, call the commander of Kirtland AFB near Albuquerque. General Koon told the base commander about the sealed orders and explained that an unusual aircraft would make a deadstick landing at Kirtland within the next half hour. The general then instructed the base commander to have air police keep everyone away from the craft and get it inside a hanger as quickly as possible.

After a half hour passed, the base commander called the Pentagon to ask where the crippled aircraft was. As he was speaking, the officer saw the U-2 touch down on the runway and remarked, "It's not a plane, it's a glider!" Even more surprised were the air police who surrounded the craft when it came to a halt. As the pilot climbed from the cockpit in his "space" suit, one air policeman remarked that the pilot looked like a man from Mars. The pilot, Jacob Kratt, later reported to Cunningham that, from the beginning of the first flameout until the landing at Albuquerque, the U-2 had covered over 900 miles, including more than 300 by gliding.[58]

Aside from this extraordinary gliding ability, however, the U-2 was a very difficult aircraft to fly. Its very light weight, which enabled it to achieve extreme altitude, also made it very fragile. The aircraft was also very sleek, and it sliced through the air with little drag. This feature was dangerous, however, because the U-2 was not built to withstand the G-forces of high speed. Pilots had to be extremely careful to keep the craft in a slightly nose-up attitude when flying at operational altitude. If the nose dropped only a degree or two into the nose-down position, the plane would gain speed at a dramatic rate, exceeding the placarded speed limit in less than a minute, at which point the aircraft would begin to come apart. Pilots, therefore, had to pay close attention to the aircraft's speed indicator because at 65,000 feet there was no physical sensation of speed, without objects close at hand for the eye to use as a reference.[59]

THREE FATAL CRASHES IN 1956

The first fatality directly connected with flying the U-2 occurred on 15 May 1956, when test pilot Wilburn S. Rose, flying article 345A, had trouble dropping his pogos, the outrigger wheels that keep the

[58] Bissell interview (S); Cunningham interview (TS Codeword); Brig. Gen. Leo A. Geary, interview by Gregory W. Pedlow, Colorado Springs, Colorado, 12 October 1988 (S).

[59] James Cherbonneaux, Carmine Vito, and Hervey Stockman (former U-2 pilots), interview by Donald E. Welzenbach, Washington, DC, May 1986 (S).

wings parallel to the ground during takeoff. Once airborne, Rose made a low-level pass over the airstrip and shook loose the lefthand pogo. When he attempted to make a righthand turn to come back over the runway to shake loose the remaining pogo, Rose stalled the U-2 and it plunged to earth, disintegrating over a wide area. Three months later, on 31 August 1956, a second fatal crash occurred during a night-flying exercise. Frank G. Grace stalled article 354 at an altitude of about 50 feet when he tried to climb too steeply at takeoff. The craft fell, cartwheeled on its left wing, and struck a power pole near the runway. More experienced U-2 pilots always cut back abruptly on the throttle as soon as the pogo sticks fell away in order to avoid such stalls.

Before the year was out, two more U-2s were destroyed in crashes, one of them fatal. On 17 September 1956, article 346 lost part of its right wing while on its takeoff ascent from Lindsey Air Force Base in Wiesbaden, Germany. The aircraft disintegrated in mid-air, killing pilot Howard Carey. The loss of article 357 on 19 December 1956 resulted from pilot hypoxia. A small leak prematurely depleted the oxygen supply and impaired Robert J. Ericson's judgment as he flew over Arizona. Because of his inability to act quickly and keep track of his aircraft's speed, the U-2 exceeded the placarded speed of 190 knots and literally disintegrated when it reached 270 knots. Ericson managed to jettison the canopy and was sucked out of the aircraft at 28,000 feet. His chute opened automatically at 15,000 feet, and he landed without injury. The aircraft was a total loss.[60]

COORDINATION OF COLLECTION REQUIREMENTS

From the very beginning of the U-2 program, it was apparent that some sort of an interagency task force or office would be needed to develop and coordinate collection requirements for the covert overhead reconnaissance effort. In a three-page memorandum to DCI Dulles on 5 November 1954 setting forth the ideas of the Technological Capabilities Panel's Project 3 on this subject, Edwin Land wrote:

> *It is recommended that... a permanent task force, including Air Force supporting elements, be set up under suitable cover to provide guidance on procurement, to consolidate requirements*

[60] U-2 Accident Reports, folders 4, 10, and 14, OSA records, job 67-B-415, box 1 (S).

and plan missions in view of priority and feasibility, to maintain the operation on a continuing basis, and to carry out the dissemination of the resulting information in a manner consistent with its special security requirements.[61]

James Q. Reber

When the U-2's development and testing approached completion, Land's recommendation was put into effect. Following a meeting with Deputy Secretary of Defense Donald Quarles and Trevor Gardner (who had been promoted from his special assistant post to become Assistant Secretary of the Air Force for Research and Development), Richard Bissell established an Ad Hoc Requirements Committee (ARC) on 1 December 1955. He then named James Q. Reber to be Intelligence Requirements Officer for the U-2 project and chairman of the ARC. Reber was already experienced in coordination with other intelligence agencies, for he had headed the Directorate of Intelligence DI Office of Intelligence Coordination for four years. The first full-scale ARC meeting took place on 1 February 1956 with representatives from the Army, Navy, and Air Force present. Attending for the CIA were representatives from the Office of Research and Reports (ORR) and the Office of Scientific Intelligence (OSI). The CIA membership later expanded to include the Office of Current Intelligence (OCI) and a representative from the Directorate of Plans. In 1957 the National Security Agency (NSA) also began sending a representative. The State Department followed suit in 1960, although it had been receiving reports from the committee all along.[62]

ARC's main task was to draw up lists of collection requirements, primarily for the U-2, but also for other means of collection. These lists prioritized targets according to their ability to meet the three major national intelligence objectives concerning the Soviet Union in the mid-1950s: long-range bombers, guided missiles, and nuclear energy. The committee issued its list of targets for the use of the entire intelligence community using all available means of collection, not just for the CIA with the U-2.[63]

---

[61] *OSA History*, chap. 1, annex 1 (TS Codeword).

[62] Minutes of the Ad Hoc Requirements Committee of 1 February 1956, Intelligence Community (IC) Staff records, COMIREX, job 33-B-121A, box 1, "ARC Minutes, 1956-1957;" Memorandum for the Joint Study Group from James Q. Reber, "Handling of Requirements for the U-2," 15 August 1960, IC Staff records, job 33-T-123A, box 10, "CHALICE (General)" (TS Codeword).

[63] Memorandum for the Joint Study Group from James Q. Reber, "Handling of Requirements for the U-2," 15 August 1960, IC Staff records, job 33-T-123A, box 10, "CHALICE (General)" (TS Codeword).

Arthur C. Lundahl

ARC gave the top priority target list to the Project Director, and the project staff's operations section then used the list to plan the flightpaths for U-2 missions. Although the requirements committee was not responsible for developing flight plans, it assisted the planners with detailed target information as required. When a flight plan was ready for submission to the President for approval, the committee drew up a detailed justification for the selection of the targets. This paper accompanied the flight plan.[64]

In developing and prioritizing lists of targets, the committee members had to take into account the varying needs and interests of their parent organizations. Thus, the CIA representatives generally emphasized strategic intelligence: aircraft and munitions factories, power-generating complexes, nuclear establishments, roads, bridges, inland waterways. In contrast, the military services usually placed a heavier emphasis on order-of-battle data. The Air Force, in particular, had a strong interest in gathering intelligence on the location of Soviet and East European airfields and radars.

Although the committee members kept the interests of their services or agencies in mind, their awareness of the vital nature of their mission kept the level of cooperation high. The group always attempted to reach a consensus before issuing its recommendations, although occasionally this was not possible and one or more agencies would add a dissent to the recommendation of the committee as a whole.[65]

PREPARATIONS TO HANDLE THE PRODUCT OF U-2 MISSIONS

On 13 December 1954, DCI Allen Dulles and his assistant, Richard Bissell, briefed Arthur C. Lundahl, the chief of CIA's Photo-Intelligence Division (PID), on Project AQUATONE. At DCI Dulles's direction, Lundahl immediately set in motion within his division a compartmented effort, known as Project EQUINE, to plan for the exploitation of overhead photography from the U-2 project. With only 13 members, the PID staff was too small to handle the expected

[64] Ibid.; James Q. Reber, interview by Donald E. Welzenbach and Gregory W. Pedlow, Washington, DC, 21 May 1987 (~~S~~).

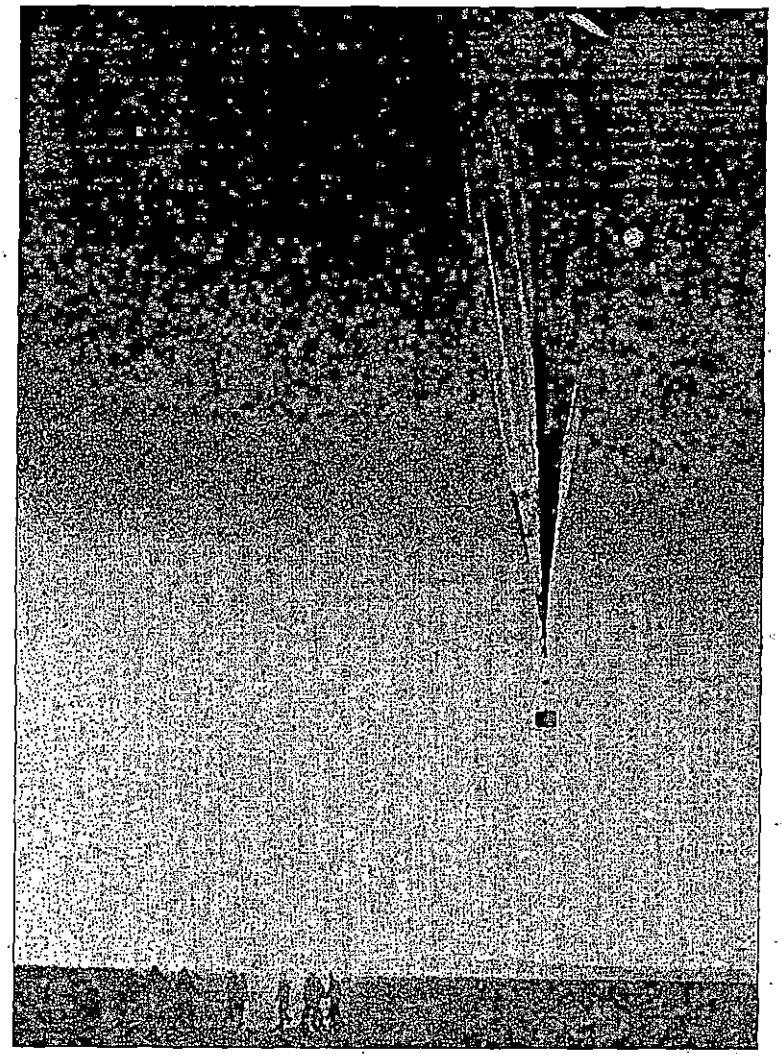
[65] Reber interview (~~S~~).

The Steuart Building, home of the Photo-Intelligence Division

flood of photographs that the U-2 would bring back, so in May 1955 the Directorate of Support (DS) authorized expanding PID to 44 persons. Soon afterward the division moved from its room in M Building to larger quarters in Que Building.

The Photo-Intelligence Division continued to expand in anticipation of large quantities of U-2 photography. Its authorized strength doubled in January 1956 when a new project known as HTAUTOMAT came into existence to exploit U-2 photography. All of the products from this project would be placed in the new control system. By the summer of 1956, the PID had moved to larger quarters in the Steuart Building at 5th Street and New York Avenue, NW. PID photointerpreters had already begun to work with U-2 photography following a series of missions in April 1956, when U-2s photographed a number of US installations that were considered analogous to high-priority Soviet installations. As a result of these preparations, PID was ready for the mass of photography that began coming when U-2 operations commenced in the summer of 1956.[66]

[66] For a more detailed history of photointerpretation in the CIA, see Urban J. Linehan, *The National Photographic Interpretation Center*, vol. 1, *Antecedents and Early Years*, Directorate of Science and Technology Historical Series NPIC-2, December 1972, pp. 171-194 (S).



Project GENETRIX balloon launch

THE IMPACT OF THE AIR FORCE PROJECT GENETRIX BALLOONS

While the Agency was making its final preparations for U-2 overflights, the Air Force started a reconnaissance project that would cause considerable protest around the world and threaten the existence of the U-2 overflight program before it even began. Project GENETRIX involved the use of camera-carrying balloons to obtain high-altitude photography of Eastern Europe, the Soviet Union, and

the People's Republic of China. This project had its origins in a RAND Corporation study from 1951. By the end of 1955, the Air Force had overcome a number of technical problems in camera design and recovery techniques and had manufactured a large number of balloons for use in the project. President Eisenhower gave his approval on 27 December 1955, and two weeks later the launches from bases in Western Europe began. By the end of February 1956, the Air Force had launched a total of 516 balloons.[67]

Project GENETRIX was much less successful than its sponsors had hoped. Once launched, the balloons were at the mercy of the prevailing winds, and many tended to drift toward southern Europe and then across the Black Sea and the desert areas of China. These balloons therefore missed the prime target areas, which lay in the higher latitudes. Large numbers of balloons did not succeed in crossing the Soviet Union and China, some because they were shot down by hostile aircraft, others because they prematurely expended their ballast supplies and descended too soon. Only 46 payloads were eventually recovered (one more than a year later and the last not until 1958) from the 516 balloons that had been launched. In four of these payloads the camera had malfunctioned, and in another eight the photography was of no intelligence value. Thus, only 34 balloons succeeded in obtaining useful photographs.[68]

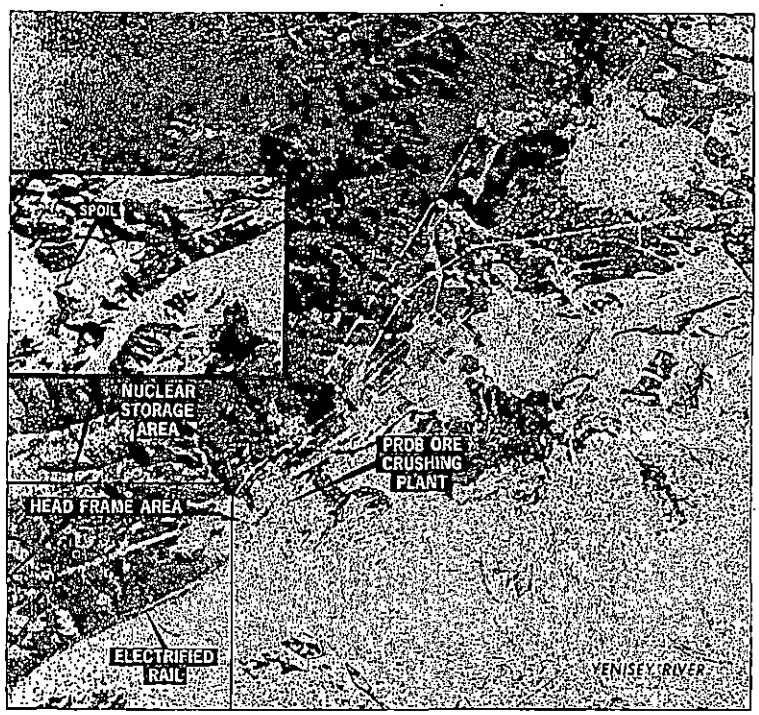
The low success rate of the Project GENETRIX balloons was not the only problem encountered; far more serious was the storm of protest and unfavorable publicity that the balloon overflights provoked. Although the Air Force had issued a cover story that the balloons were being used for weather research connected with the International Geophysical Year, East European nations protested strongly to the United States and to international aviation authorities, claiming that the balloons endangered civilian aircraft. The Soviet Union sent strongly worded protest notes to the United States and the nations from which the balloons had been launched. The Soviets also collected numerous polyethelene gasbags, camera payloads, and transmitters from GENETRIX balloons and put them on display in Moscow for the world press.[69]

---

[67] P. G. Strong, Attachment to Memorandum for DCI Dulles, "Project GENETRIX Summary," 15 February 1956 ~~(S)~~.

[68] Final Report, Project 119L, 1st Air Division (Meteorological Survey) Strategic Air Command, 5 March 1956, D-582, General Summary ~~(S~~, declassified 1979).

[69] *New York Times*, 10 February 1956, p. 1; *Omaha World Herald*, 11 February 1956, p. 1.



Photograph of Dodonovo Atomic Energy complex taken by a Project GENETRIX balloon

All of this publicity and protest led President Eisenhower to conclude that "the balloons gave more legitimate grounds for irritation than could be matched by the good obtained from them," and he ordered the project halted. On 7 February 1956 Secretary of State Dulles informed the Soviet Union that no more "weather research" balloons would be released, but he did not offer an apology for the overflights.[70]

Despite the furor caused by GENETRIX, Air Force Chief of Staff Twining proposed yet another balloon project only five weeks later, in mid-March 1956. This project would employ even higher flying balloons than GENETRIX and would be ready in 18 months. President Eisenhower informed the Air Force, however, that he was "not interested in any more balloons."[71]

[70] Andrew J. Goodpaster, Memorandum for the Record, "10 February 1956 Conference of Joint Chiefs of Staff with the President," WHOSS, Alpha, DDEL ~~CPS~~, declassified 1980); Stephen E. Ambrose, *Eisenhower: The President* vol. 2 (New York: Simon and Schuster, 1984), p. 310.

[71] Quoted in Ambrose, *Eisenhower: The President*, p. 310.

Although the photo intelligence gained from Project GENETRIX was limited in quantity, it was still some of the best and most complete photography obtained of the Soviet Union since World War II. It was referred to as "pioneer" photography because it provided a baseline for all future overhead photography. Even innocuous photos of such things as forests and streams proved valuable in later years when U-2 and satellite photography revealed construction activity.

Of still greater importance to the U-2 program, however, was the data that US and NATO radars obtained as they tracked the paths of the balloons—whose average altitude was 45,800 feet—over the Soviet Bloc. This data provided the most accurate record to date of high-altitude wind currents, knowledge that meteorologists were later able to put to use to determine optimum flightpaths for U-2 flights.

One completely fortuitous development from Project GENETRIX had nothing to do with the cameras but involved a steel bar. This bar served a dual purpose: the rigging of the huge polyethylene gasbag was secured to the top of the bar and the camera-payload and automatic-ballasting equipment was attached to the bottom. By sheer chance, the length of the bar—91 centimeters—corresponded to the wavelength of the radio frequency used by a Soviet radar known by its NATO designator as TOKEN. This was an S-band radar used by Soviet forces for early warning and ground-controlled intercept. The bar on the GENETRIX balloons resonated when struck by TOKEN radar pulses, making it possible for radar operators at US and NATO installations on the periphery of the Soviet Union to locate a number of previously unknown TOKEN radars.

These radar findings, coupled with other intercepts made during the balloon flights, provided extensive data on Warsaw Pact radar networks, radar sets, and ground-controlled interception techniques. Analysis of these intercepts revealed the altitude capabilities and tracking accuracy of radars, the methods used by Warsaw Pact nations to notify each other of the balloons' passage (handing off), and the altitudes at which Soviet aircraft could intercept the balloons. All of this information could be directly applied to future U-2 missions.[72]

[72] Final Report, Project 119L, 1st Air Division (Meteorological Survey) Strategic Air Command, 5 March 1956, D-582, General Summary (~~TS~~, declassified 1979).

These positive results from Project GENETRIX did not outweigh the political liabilities of the international protests. CIA officials became concerned that the ill will generated by balloon overflights could sour the Eisenhower administration on all overflights, including those by the U-2, which was just about ready for deployment. Therefore, DDCI Cabell wrote to Air Force Chief of Staff Twining in February 1956 to warn against further balloon flights because of the "additional political pressures being generated against all balloon operations and overflights, thus increasing the difficulties of policy decisions which would permit such operations in the future." [73]

In addition to its concern for the future of the U-2 program, the Agency feared that President Eisenhower's anger at balloon overflights might result in the curtailment of the balloon program that the Free Europe Committee—a covert Agency operation based in West Germany—used to release propaganda pamphlets over Eastern Europe.

AQUATONE BRIEFINGS FOR SELECTED MEMBERS OF CONGRESS

Although knowledge of the U-2 project was a closely guarded secret within both the Agency and the Eisenhower administration, DCI Dulles decided that a few key members of Congress should be told about the project. On 24 February 1956, Dulles met with Senators Leverett Saltonstall and Richard B. Russell, the ranking members of the Senate Armed Services Committee and its subcommittee on the CIA. He shared with them the details of Project AQUATONE and then asked their opinion on whether some members of the House of Representatives should also be informed. As a result of the senators' recommendation that the senior members of the House Appropriations Committee should be briefed, Dulles later met with its ranking members, Representatives John Taber and Clarence Cannon. Official Congressional knowledge of the U-2 project remained confined to this small group for the next four years. The House Armed Services Committee and its CIA subcommittee did not receive a CIA briefing on the U-2 project until after the loss of Francis Gary Powers's U-2 over the Soviet Union in May 1960.[74]

[73] Philip G. Strong, Attachment to Memorandum for DCI Dulles, "Project GENETRIX Summary," 15 February 1956, OSI records (S).

[74] John S. Warner, Legislative Counsel, Memorandum for the Record, "AQUATONE Briefings," 18 November 1957, Office of Congressional Affairs records, job 61-357, box 2 (S); Warner interview (S).

THE U-2 COVER STORY

In February 1956, while the controversy over balloon flights was still raging and the U-2 was completing its final airworthiness tests, Richard Bissell and his staff began working on a cover story for overseas operations. It was important to have a plausible reason for deploying such an unusual looking plane, whose glider wings and odd landing gear were certain to arouse curiosity.

Bissell decided that the best cover for the deployment of the U-2 was an ostensible mission of high-altitude weather research by the National Advisory Committee on Aeronautics (NACA). Such a cover story, however, needed the approval of all concerned: Air Force intelligence, the Air Weather Service, the Third Air Force, the Seventh Air Division, the SAC U-2 project officer, the Air Force Headquarters project officer, and NACA's top official, Dr. Hugh Dryden. Moreover, the CIA Scientific Advisory Committee was also consulted about the cover plan.

Senior CIA officials and the other agencies involved in providing cover for the U-2 approved the final version of the overall cover story at the end of March 1956. The project staff then began working on contingency plans for the loss of a U-2 over hostile territory. Bissell advised the project's cover officer to "produce a document which sets forth all actions to be taken . . . not only press releases and the public *line* to be taken, but also the suspension of operations and at least an indication of the diplomatic action. . . .We should at least make the attempt in this case to be prepared for the worst in a really orderly fashion." The cover officer then prepared emergency procedures based on the overall weather research cover story, and Bissell approved these plans. There was one final high-level look at the cover story on 21 June 1956, the day after the first U-2 mission over Eastern Europe, when Bissell met with General Goodpaster, James Killian, and Edwin Land to discuss the pending overflights of the Soviet Union, including the proposed emergency procedures. Killian and Land disagreed with Bissell's concept and made a much bolder and more forthright proposal: in the event of the loss of a U-2 over hostile territory, the United States should not try to deny responsibility but should state that overflights were being conducted "to guard against surprise attack." This proposal was put aside for further thought (which it never received), and Bissell's weather research cover remained the basis for statements to be made after a loss. The project staff then went on to prepare a number of different statements to be

used in various scenarios, including one in which the pilot was captured. Even in such a case, however, the proposed policy was for the United States to stick to the weather research cover story, a course of action that would prove disastrous in May 1960.[75]

[75] *OSA History*, chap. 8, pp. 30-35; chap. 11, annex 73 (~~TS~~ Codeword).

U-2 Operations in the Soviet Bloc and Middle East, 1956-1968

By January 1956, everyone working on Project AQUATONE could see that the U-2 was nearing the time for operational deployment. During tests the aircraft had met all the criteria established in late 1954. Its range of 2,950 miles was sufficient to overfly continents, its altitude of 72,000 feet was beyond the reach of all known antiaircraft weapons and interceptor aircraft, and its camera lenses were the finest available.

Because the main targets for the U-2 lay behind the Iron Curtain, Bissell and his staff began looking for operational bases in Europe. The United Kingdom, America's closest ally, seemed the logical choice for U-2 bases, and, on 10 January 1956, Bissell flew to London to discuss the matter with Royal Air Force (RAF) and MI-6 officials. Their initial response was favorable, but they told Bissell that the proposal needed approval at a much higher level.

Bissell reported his findings to DCI Dulles, who promptly arranged to meet with Foreign Secretary Selwyn Lloyd in London to explore the possibility of winning the British Government's approval for the project. Dulles presented his case to Lloyd on 2 February, and, by early March, Prime Minister Anthony Eden approved the basing of U-2s in the United Kingdom. The U-2s were to use Lakenheath AFB, an RAF base also used by the USAF Strategic Air Command (SAC).[1]

[1] *OSA History*, chap. 11, pp. 10-15 (~~TS~~ Codeword).

THE DEPLOYMENT OF DETACHMENT A TO LAKENHEATH

The first Agency U-2 detachment, consisting of four aircraft and pilots, was known publicly as the 1st Weather Reconnaissance Squadron, Provisional (WRSP-1). The "provisional" designation gave the U-2 detachments greater security because provisional Air Force units did not have to report to higher headquarters. WRSP-1, known within the Agency as Detachment A, began deploying to the United Kingdom on 29 April 1956. By 4 May, all of the detachment's personnel and equipment, including four aircraft, had arrived at Lakenheath.[2]

Shortly after deployment, on 7 May, the National Advisory Committee on Aeronautics (NACA) released an unclassified U-2 cover story stating that a Lockheed-developed aircraft would be flown by the USAF Air Weather Service to study such high-altitude phenomena as the jet stream, convective clouds, temperature and wind structures at jet-stream levels, and cosmic-ray effects up to 55,000 feet.[3]

Before overflights could begin from Lakenheath, however, several incidents occurred that dampened Prime Minister Eden's interest in having the U-2s on British territory. In mid-April 1956, a Soviet naval squadron brought Soviet leaders Nikita Khrushchev and Nikolai Bulganin on an official visit to the United Kingdom. Although the ships were docked in Portsmouth Harbor, a British counterintelligence operative and underwater expert, retired Royal Navy Commander Lionel Crabb, apparently undertook a mission to examine the hulls of these vessels but vanished in the process. His headless body was later found washed up on a beach. This so-called Frogman Incident caused an uproar in Parliament and a protest from Moscow that soured relations between the United Kingdom and the Soviet Union. To avoid further deterioration of Anglo-Soviet relations, the Prime Minister wrote to President Eisenhower on 16 May asking that overflights be postponed. Only two days later, a U-2 on a training flight from Lakenheath inadvertently penetrated the British radar network, causing RAF fighters to scramble. Afterward the Air Ministry made a public announcement that a special NACA aircraft was conducting high-altitude research in the United Kingdom. At about the

[2] Ibid., pp. 17-18 (~~TS~~ Codeword).

[3] Press Release of 7 May 1956 (U) in *OSA History*, chap. 7, annex 60 (~~TS~~ Codeword).

same time, Richard Bissell learned that the State Department had told Prime Minister Eden that only one U-2 was based at Lakenheath, when in reality there were four.[4]

THE MOVE TO WIESBADEN

To avoid arousing further reaction in the United Kingdom and to begin the program of U-2 overflights beyond the Iron Curtain without further delay, Bissell moved Detachment A on 11 June 1956 to Wiesbaden, one of the busiest airfields in West Germany, without notifying West German authorities. The detachment commander, Col. Frederick McCoy, was disappointed in his hope that the redeployment of the U-2s could be accomplished without drawing undue attention. The strange-looking planes, with bicycle-type wheels and wings so long they touched the ground after landing, aroused considerable interest. Wiesbaden was to be only a temporary home for Detachment A; the Air Force began preparing Giebelstadt near the East German border for use by the U-2s. Giebelstadt was an old World War II airbase that had been one of the launching sites for the GENETRIX balloons.[5]

Soon after the four U-2s arrived in Wiesbaden, they were refitted with the more powerful J57/P-31 engines. The new engines were better suited for operations behind the Iron Curtain because they were less likely to suffer flameouts than the earlier model. Once the new engines were installed, the aircraft received the designation U-2B.[6]

Bissell was anxious to get the overflights started by late June because SAC weather experts had predicted that the best weather for photographing the Soviet Union would be between 20 June and 10 July. Bissell, however, had not yet received final authorization from President Eisenhower to begin overflights of the Soviet Union. On 28 May 1956, when DCI Allen Dulles met with the President to discuss the U-2's readiness for operations, Eisenhower still made no decision on overflights. Three days later Dulles and Air Force Chief of Staff

[4] Christopher Andrew, *Her Majesty's Secret Service: The Making of the British Intelligence Community* (New York, 1986), pp. 495-496; Beschloss, *Mayday*, p. 116; *OSA History*, chap. 11, pp. 18-21 (~~TS~~ Codeword).

[5] *OSA History*, chap. 11, pp. 21-23 (~~TS~~ Codeword).

[6] Ibid., pp. 23, 26 (~~TS~~ Codeword).

Nathan Twining prepared a paper for the President outlining "AQUATONE Operational Plans." In the meantime, President Eisenhower had entered Walter Reed Hospital for tests for an abdominal ailment that turned out to be ileitis, requiring an operation. During his recovery from surgery, Eisenhower would make his final decision on the overflight program.[7]

PRESIDENT EISENHOWER'S ATTITUDE TOWARD OVERFLIGHTS

The President had mixed feelings about overflights of the Soviet Union. Aware that they could provide extremely valuable intelligence about Soviet capabilities, he, nevertheless, remained deeply concerned that such flights brought with them the risk of starting a war. From the very beginning of the U-2 program, President Eisenhower had worked to minimize the possibility that overflights could lead to hostilities. He had always insisted that overflights by military aircraft were too provocative, and in 1954 he had therefore supported the Land committee's proposal for an unarmed civilian aircraft instead of the military reconnaissance planes favored by the Air Force. For the same reason, Eisenhower had resisted attempts by the Air Force to take the U-2 program away from the CIA in 1955.

In fact, the President's desire to avoid secret reconnaissance missions over the Soviet Union, with all their risks, led him to make his famous "Open Skies" proposal in the summer of 1955, when the U-2 was still under development but making good progress. At the Geneva summit conference on 21 July 1955, President Eisenhower offered to provide airfields and other facilities in the United States for the Soviet Union to conduct aerial photography of all US military installations if the Soviet Union would provide the United States with similar facilities in Russia. Not surprisingly, Soviet leader Nikita Khrushchev almost immediately rejected Eisenhower's offer. Although the President had hoped that the Soviet Union would accept his proposal, he was prepared for rejection. While Open Skies was still being considered, Eisenhower had stated, "I'll give it one shot. Then if they don't accept it, we'll fly the U-2."[8]

[7] Ibid., pp. 23-25 and annex 73, "AQUATONE Operational Plans," 31 May 1956 (~~TS~~ Codeword).

[8] Quoted in Beschloss, *Mayday*, p. 105.

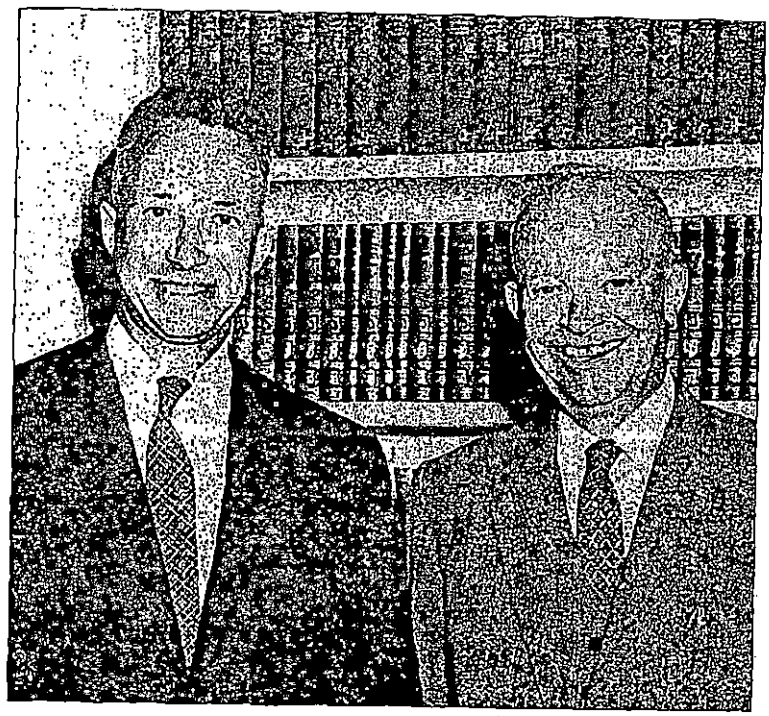
Even though President Eisenhower had approved every stage of the U-2's development, knowing full well that the aircraft was being built to fly over the Soviet Union, the actual decision to authorize such flights was very difficult for him. He remained concerned that overflights could poison relations with the Soviet Union and might even lead to hostilities. One argument that helped overcome the President's reluctance was the CIA's longstanding contention that U-2 flights might actually go undetected because Soviet radars would not be able to track aircraft at such high altitudes. This belief was based on a 1952 study of Soviet World War II–vintage radars and on 1955 tests using US radars, which—unknown to US officials—were not as effective as Soviet radars against high-altitude targets. Shortly before U-2 operations began, however, the CIA's Office of Scientific Intelligence (OSI) conducted a vulnerability study of the U-2 that was published on 28 May 1956. The study's conclusion was that "Maximum Soviet radar detection ranges against the Project aircraft at elevation in excess of 55,000 feet would vary from 20 to 150 miles.... In our opinion, detection can therefore be assumed." The OSI study added, however, "It is doubtful that the Soviets can achieve consistent tracking of the Project vehicle."[9] Completed just three weeks before the initation of overflights, this study seems to have had little impact on the thinking of the top project officials. They continued to believe that the Soviets would not be able to track the U-2 and might even fail to detect it, except for possible vague indications.[10]

Soviet radars were not President Eisenhower's only concern. Also fearing that a malfunction might cause a U-2 to crash inside the Soviet Union, he asked Allen Dulles what the consequences would be. The President's staff secretary, Col. Andrew J. Goodpaster, who was present at virtually all White House meetings on the U-2 project and served as the President's intermediary to the CIA on this issue, later recalled:

> *Allen's approach was that we were unlikely to lose one. If we did lose one, the pilot would not survive.... We were told—and it was part of our understanding of the situation—that it was almost certain that the plane would disintegrate and that we could*

[9] *OSA History*, chap. 11, p. 31 (~~TS~~ Codeword). For the belief that the U-2 might go undetected see the Leghorn interview and Dwight D. Eisenhower, *Waging Peace, 1956-1961* (New York, 1965), p. 41.

[10] Richard M. Bissell, Jr., interview by Gregory W. Pedlow, tape recording, Farmington, Connecticut, 28 October 1988 (~~S~~).



Colonel Goodpaster with President Eisenhower

take it as a certainty that no pilot would survive . . . and that although they would know where the plane came from, it would be difficult to prove it in any convincing way.[11]

CIA assurances that the U-2 would probably not be detected, and that a crashed U-2 could not be traced back to the United States, helped overcome the President's worries about overflights. The most important reason why President Eisenhower decided to send reconnaissance aircraft over the Soviet Union, however, was the urgent need for accurate intelligence to confirm or disprove claims of Soviet advances in long-range bombers and missiles. The initial sighting of the new Soviet Bison bomber in the spring of 1954 had been followed by reported sightings of more than 30 of these bombers in the spring and summer of 1955 (in reality these were sightings of the same group of 10 aircraft that circled around out of sight and made several passes during a Soviet air show). Soon members of Congress were calling for investigations into the relative strength of the US and

[11] Quoted in Beschloss, *Mayday*, p. 118.

Soviet Air Forces.[12] Early in 1956, concern about a possible Soviet advantage in long-range bombers grew as Air Force Chief of Staff Twining informed the Senate Armed Services Committee that the Soviet Union already had more Bisons than the United States had B-52s and that the Soviets would be able to "maintain this advantage for some time if they keep on the production curve we are now predicting." [13] By May 1956, reporting on the growing Soviet air strength was no longer confined to aviation journals; *U.S. News and World Report*, for example, featured articles headlined "Can Soviets Take the Air Lead?" and "Is U.S. Really Losing in the Air?" [14]

Alongside fear of possible Soviet superiority in long-range bombers came a new potential threat: Soviet progress in guided missile research. Trevor Gardner, Air Force Assistant Secretary for Research and Development, warned in September 1955 that "the most complex and baffling technological mystery today is not the Russian capability in aircraft and nuclear weapons but rather what the Soviet progress has been in the field of guided missiles." [15] On 30 January 1956, *Time* magazine made the guided missile its cover story. The article began by describing a hypothetical crisis set in 1962 in which the United States suffered a humiliating defeat because it had lagged behind the Soviet Union in guided missile development.[16] Just two weeks after this story appeared, the Soviets successfully tested a missile with a range of 900 miles, and President Eisenhower admitted at a press conference that the Soviet Union might be ahead of the United States in some areas of the missile field. Administration critic Senator Stuart Symington then claimed, "The facts are that our missile development may be ahead in the short-range area, but their missile development is ahead in the area that counts by far the most—the

[12] Robert Hotz, "Russian Jet Airpower Gains Fast on US," *Aviation Week*, 23 May 1955, pp. 12-15; "Aviation Week Story Spurs Debate on US, Red Airpower Positions," *Aviation Week*, 30 May 1955, pp. 13-14.

[13] Claude Witze, "Russians Outpacing US in Air Quality, Twining Warns Congress," *Aviation Week*, 27 February 1956, pp. 26-28; Robert Hotz, "Russian Air Force Now Gaining in Quality," *Aviation Week*, 12 March 1956, p. 286.

[14] "Can Soviets Take the Air Lead? What LeMay, Wilson, Ike Say," *US News and World Report*, 11 May 1956, pp. 108-114; "Is U.S. Really Losing in the Air?" *US News and World Report*, 18 May 1956, pp. 25-27.

[15] William Coughlin, "Gardner Defends Greater R&D Spending," *Aviation Week*, 26 September 1955, p. 14.

[16] "Missiles Away," *Time*, 30 January 1956, pp. 52-55.

long-range area." [17] Fears of Soviet missile progress increased when Nikita Khrushchev stated on 23 April 1956, "I am quite sure that we shall have very soon a guided missile with a hydrogen-bomb warhead which could hit any point in the world." [18]

Faced with growing Congressional and public anxiety over Soviet offensive capabilties, President Eisenhower approved the proposed overflight program. Colonel Goodpaster relayed this decision to Bissell, Land, and Killian at a meeting on 21 June. The President nevertheless maintained tight control over the program and authorized only 10 days of overflights when operations over the Soviet Union were ready to start in early July 1956.[19]

FIRST OVERFLIGHTS OF EASTERN EUROPE

The CIA initiated U-2 flights over hostile territory even before the President granted final approval for overflights of the Soviet Union. After consulting with the Commander of US Air Force Europe, Richard Bissell used existing Presidential permission for Air Force overflights of the Soviet Union's East European satellites as his authority to plan a mission over Poland and East Germany. Bissell had informed the President of his intention to conduct such missions in the "AQUATONE Operational Plan" submitted on 31 May.

The first operational use of a U-2 took place on Wednesday, 20 June 1956. Carl K. Overstreet flew a U-2 equipped with an A-2 camera over Poland and East Germany. At the end of the mission, Detachment A immediately rushed the exposed film to the United States for processing. The developed film arrived at the Photo-Intelligence Division (PID) on 22 June 1956. PID personnel considered the pictures obtained by mission 2003 to be of good quality.[20]

---

[17] Robert Hotz, "Firing of 900-Mile Russian Missile Spurs US Changes," *Aviation Week*, 20 February 1956, p. 27.

[18] "Is Russia Really Ahead in Missile Race?," *US News and World Report*, 4 May 1956, p. 34.

[19] *OSA History*, chap. 11, pp. 27-29 (TS Codeword); A. J. Goodpaster, Memorandum for the Record, 21 June 1956, WHOSS, Alpha, DDEL (~~TS~~).

[20] *OSA History*, chap. 11, p. 27 (~~TS~~ Codeword); Mission folder 2003 (20 June 1956), OSA records, job 67-B-328, box 7 (~~TS~~ Codeword).

Following the success of this first mission, Bissell was eager to begin overflights of the Soviet Union. But even after the President granted his approval on 21 June, such missions could not yet take place for two reasons. First, President Eisenhower had agreed with a CIA and State Department recommendation that West German Chancellor Konrad Adenauer be informed in advance of US plans to overfly the Soviet Union from bases in Germany (in keeping with existing policies Adenauer was not informed about overflights of Eastern Europe). Second, Soviet party chief Nikita Khrushchev had invited representatives of the US Air Force to the Moscow Air Show, which opened on 23 June 1956. Led by Air Force Chief of Staff Nathan F. Twining, the delegation would be in the Soviet Union for a week, and General Twining requested that no overflights of the Soviet Union be staged until the Air Force delegation had left.[21]

Both of these restrictions on overflights of the Soviet Union were cleared up by the end of June. Accompanied by General Cabell and the Chief of Station for Germany, Tracy Barnes, Bissell briefed Chancellor Adenauer and his trusted adviser, State Secretary Hans Globke, on 27 June. Adenauer enthusiastically endorsed the project.[22]

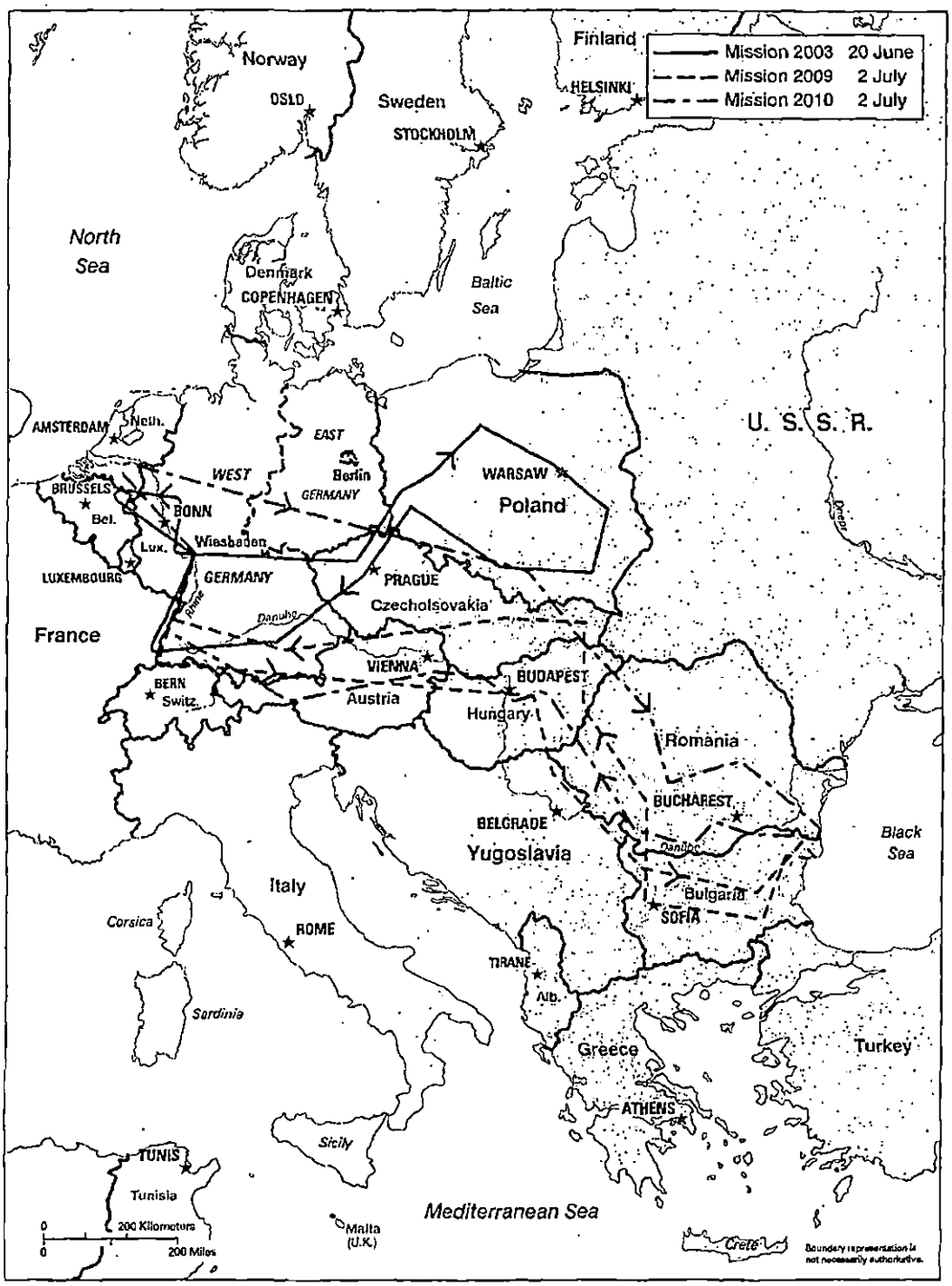
A few days later the Air Force delegation returned from Moscow, but now unfavorable weather prevented the start of operations against the Soviet Union.

While waiting for the clouds over the Soviet Union to clear, Detachment A carried out two more overflights of Eastern Europe on 2 July 1956: mission 2009 over Czechoslovakia, Hungary, and Bulgaria; and mission 2010 over East Germany, Poland, Hungary, and Romania. That afternoon Bissell and DDCI Cabell gave President Eisenhower a detailed briefing on the first U-2 overflight, which the President found "very interesting, very positive." Eisenhower was anxious to know, however, whether radars had tracked the aircraft. Bissell replied that, although East European radars had picked up the 20 June flight, radar operators had misread the altitude as only 42,000 feet. He added that the Agency was awaiting reports on that morning's flights to see if they, too, had been detected. Noting that the U-2

[21] Nathan F. Twining, *Neither Liberty nor Safety* (New York: Holt, Rinehart & Winston, 1966), pp. 259-260; *OSA History*, chap. 11, p. 27 (~~TS~~ Codeword).

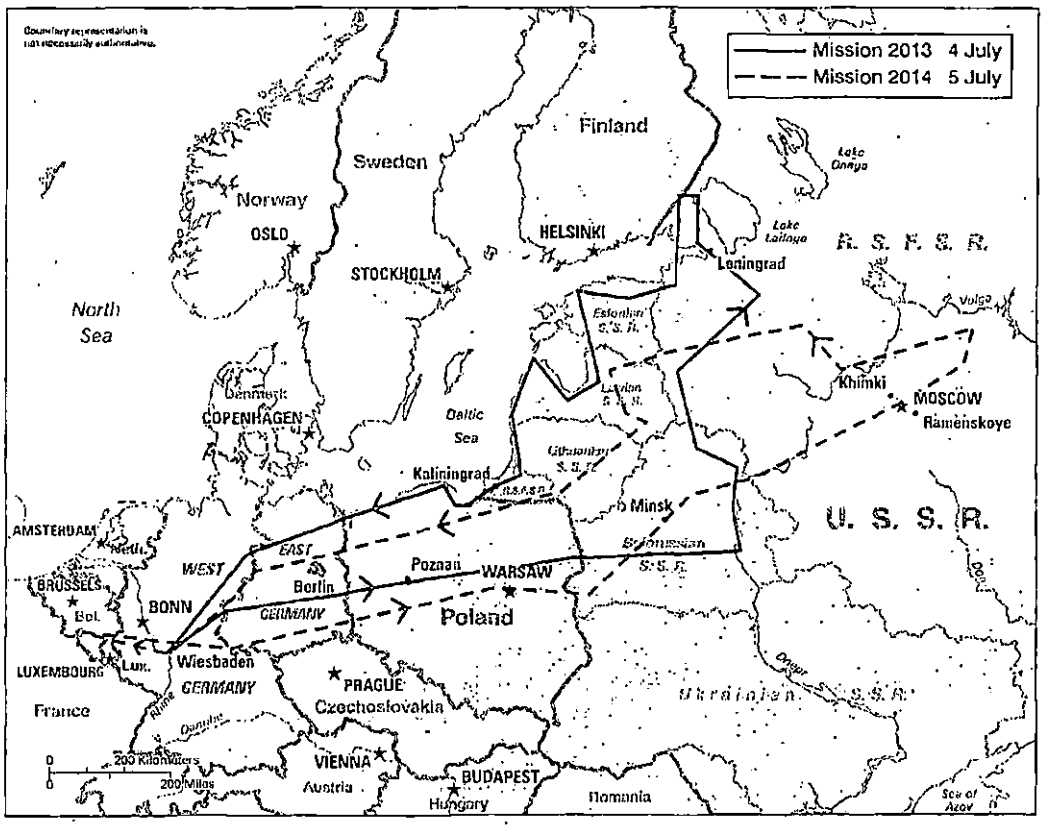
[22] *OSA History*, chap. 11, p. 28 (~~TS~~ Codeword).

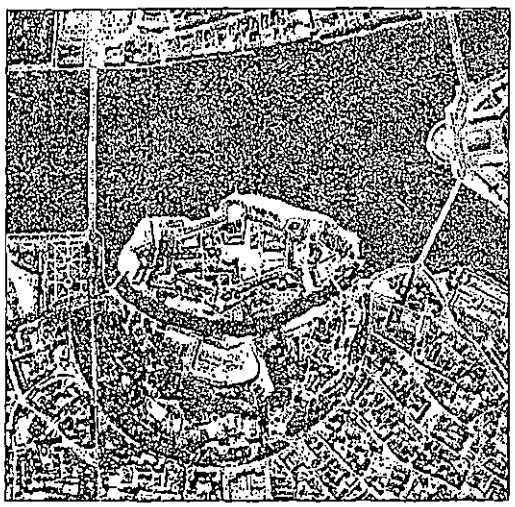
First Overflights, 20 June - 5 July 1956



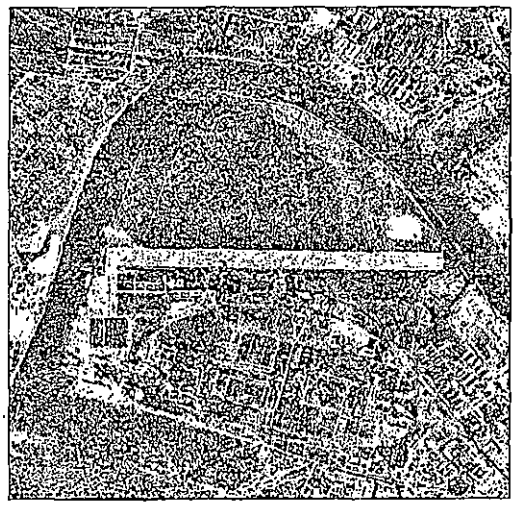
~~Secret~~ NOFORN

724742 (801373) 4-92





U-2 photography of Leningrad, 4 July 1956



Fili Airframe Plant, Moscow, 5 July 1956

detachment had four aircraft working and could average up to two flights per day, Bissell told the President that the crews were "ready and eager to go in beyond the satellites" and overfly the center of the Soviet Union.[23]

Eisenhower replied that he thought it "urgent" to know whether the recent flights had been tracked by hostile radars. The President was obviously concerned that CIA estimates that the U-2 could fly virtually undetected were proving false. One of the reasons why he had approved the overflight program was the CIA's assurance that the Soviet Union would remain unaware of the flights or—at the very worst—receive only occasional, vague indications.

FIRST U-2 FLIGHTS OVER THE SOVIET UNION

The question of how well the Soviets could track U-2 flights had not yet been settled when the first overflights of the Soviet Union took place. On Wednesday, 4 July 1956, the U-2 known as Article 347 began the first flight over the Soviet Union. Final authorization for mission 2013 had come shortly before takeoff. Late on the evening of 3 July, Bissell went to project headquarters in the Matomic Building to give the "Go" or "No go" decision. Although the President had approved the overflight, the final decision to start a mission depended on a number of factors, especially the weather over the target area and at the takeoff and landing sites. Bissell made the decision just before midnight Washington time, which was six o'clock in the morning in Wiesbaden. This pattern of last-minute approvals continued for the duration of the U-2 overflight program.[24]

When Wiesbaden received the "Go" signal, a U-2 equipped with an A-2 camera and flown by pilot Hervey Stockman took off on a course that took it over Poznan, Poland, where riots had occurred on 28-30 June. After Poznan, Stockman headed for Belorussia, where he turned north to Leningrad. The last leg of the mission took the U-2 over the Soviet Baltic states before returning to Wiesbaden. The main target of this mission was the naval shipyards in Leningrad, center of

[23] Andrew J. Goodpaster's handwritten notes on 2 July 1956 meeting, WHOSS, Alpha, DDEL (~~TS~~).

[24] Bissell interview by Welzenbach (S); Cunningham interview (~~TS~~ Codeword).

the Soviet Union's submarine construction program. Mission 2013's route also overflew a number of major military airfields to make an inventory of the new Bison jet-engine heavy bomber.[25]

The second overflight, on the following day, continued the search for Bison bombers. Pilot Carmine Vito's route was similar but somewhat to the south of Stockman's and also flew farther east, more than 200 kilometers past Moscow. Although the Soviet capital was almost completely hidden by clouds, the A-2 camera with haze filters took some usable photographs of the city. These turned out to be the only U-2 photographs of Moscow because no other mission was sent over the Soviet capital. Among the key targets photographed during mission 2014 were the Fili airframe plant, where the Bison was being built; the bomber arsenal at Ramenskoye, where the Bisons were tested; the Kaliningrad missile plant; and the Khimki rocket-engine plant.[26]

When Allen Dulles returned to work on Thursday, 5 July 1956, he asked Bissell if any overflights had taken place during the Independence Day holiday. One had been made on the fourth and another just that morning, Bissell replied. (Because of the six-hour time difference, the 5 July flight was safely back in Wiesbaden by the time Dulles spoke to Bissell.) When Dulles asked the routes of these missions, Bissell told him that they had overflown both Moscow and Leningrad. "Oh my Lord," Dulles exclaimed, "do you think that was wise the first time?" "Allen," Bissell replied, "the first is the safest."[27]

President Eisenhower also wanted to know the results of the 4 and 5 July flights, but his principal concern was whether there had been any indication that either flight had been discovered or tracked by radar. Eisenhower told Colonel Goodpaster "to advise Mr. Allen

[25] Urban J. Linehan, *National Photographic Interpretation Center: The Years of Project HTAUTOMAT, 1956-1958*, Directorate of Science and Technology Historical Series NPIC-3, December 1974, 6 vols. (hereafter cited as *NPIC History*), vol. 1, p. 20 (S); Mission folder 2013 (4 July 1956), OSA records, job 67-B-328, box 7 (~~PS~~ Codeword). Note on mission numbers: each proposed mission received a number, but not all of these missions were flown.

[26] *NPIC History*, vol. 1, p. 21 (S); Mission folder 2014 (5 July 1956), OSA records, job 67-B-328, box 7 (~~PS~~ Codeword).

[27] Bissell interview by Welzenbach ~~(S)~~.

Dulles that if we obtain any information or warning that any of the flights has been discovered or tracked, the operation should be suspended." Goodpaster called both Dulles and Bissell and was told that reports on tracking or attempted interception of the U-2s would not be available for another 36 hours. Later that day the two CIA officials met with Goodpaster to ask if flights could continue in the meantime. Goodpaster replied that his understanding of the President's directive was that the operation should continue "at the maximum rate until the first evidence of tracking was received." [28]

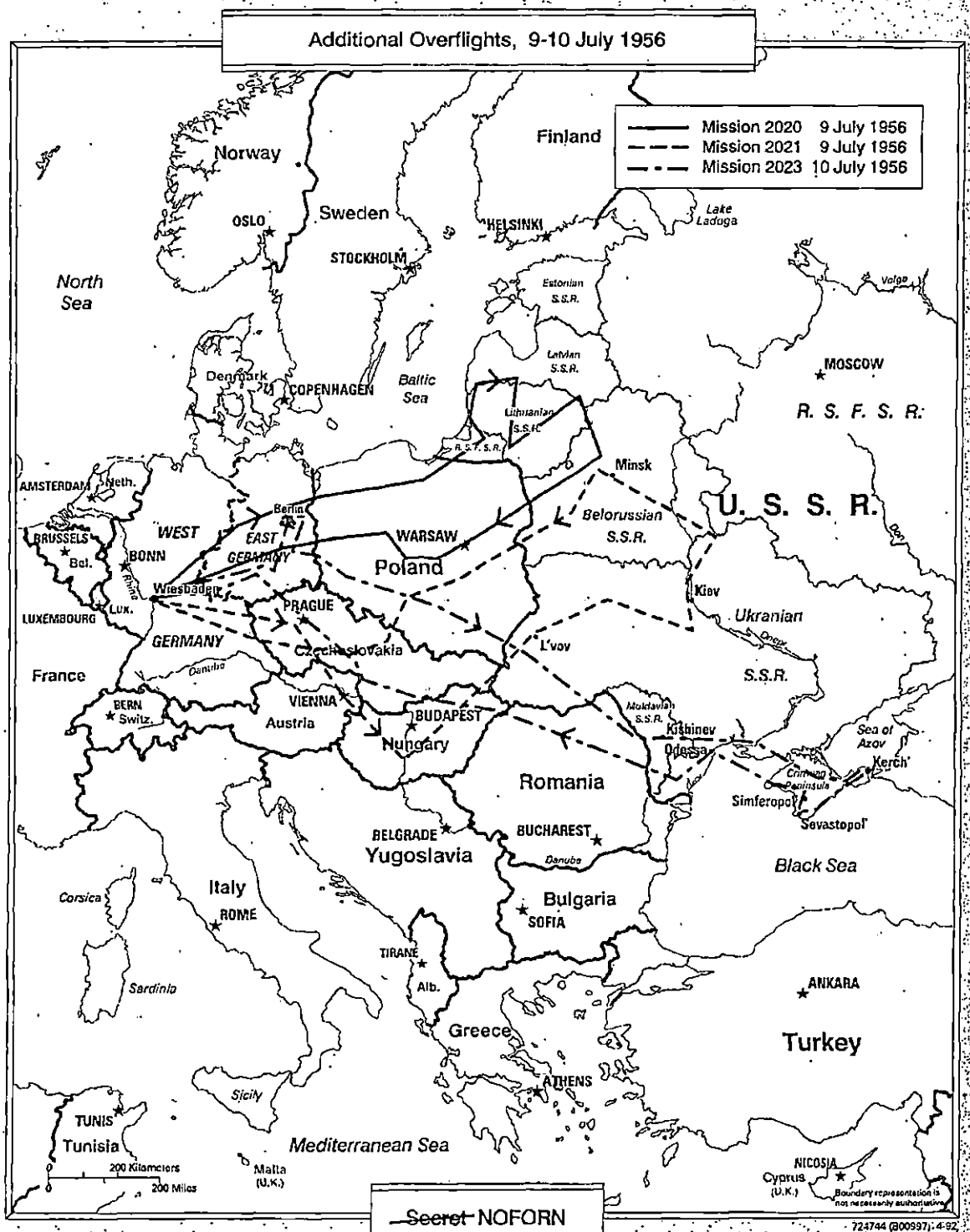
Although President Eisenhower had originally spoken of suspending the overflights if they were "discovered or tracked," his main concern was to learn if the Soviets could track U-2 missions, meaning that they could follow the flight on their radar screens for most or all of the missions and thus have numerous opportunities to attempt interception. Certainly the President hoped that U-2 flights could not even be detected, but reports received on the 20 June overflight of Eastern Europe had already indicated that this goal was unrealistic. The President's emphasis therefore shifted to tracking. If the Soviets could successfully track U-2 missions, he wanted the overflights halted.[29] Reports on Soviet radar coverage of the first two overflights of the Soviet Union became available on 6 July. These reports showed that, although the Soviets did detect the aircraft and made several very unsuccessful attempts at interception, they could not track U-2s consistently. Interestingly, the Soviet radar coverage was weakest around the most important targets, Moscow and Leningrad, and the Soviets did not realize that U-2s had overflown these two cities.[30]

Detachment A carried out three more overflights of the Soviet Union during the 10-day period authorized by the President. Two of the missions (2020 and 2021) took place on a single day, 9 July 1956. They covered much of Eastern Europe, and the Ukraine and Belorussia in the Soviet Union. Unfortunately, a broken camera

[28] Andrew J. Goodpaster, Memorandum for the Record, 5 July 1956, WHOSS, Alpha, DDEL (~~TS~~).

[29] Gen. Andrew J. Goodpaster, interview by Donald E. Welzenbach and Gregory W. Pedlow, Washington, DC, 8 July 1987 (~~S~~).

[30] Mission folders 2013 (4 July 1956) and 2014 (5 July 1956), OSA records, job 67-B-328, box 7 (~~TS~~ Codeword).


Additional Overflights, 9-10 July 1956
Mission 2020 9 July 1956
Mission 2021 9 July 1956
Mission 2023 10 July 1956
Norway
OSLO
Sweden
STOCKHOLM
Finland
HELSINKI
Lake Ladoga
North Sea
Estonian S.S.R.
Volga
Latvian S.S.R.
Denmark
COPENHAGEN
Baltic Sea
Lithuanian S.S.R.
R.S.F.S.R.
MOSCOW
R. S. F. S. R.
Minsk
AMSTERDAM
Neth.
Berlin
WEST GERMANY
EAST GERMANY
WARSAW
Belorussian S.S.R.
U. S. S. R.
BRUSSELS
Bel.
BONN
Rhine
Wiesbaden
Poland
Kiev
LUXEMBOURG
Lux.
PRAGUE
Ukranian S.S.R.
GERMANY
Czechoslovakia
L'vov
France
Danube
VIENNA
Austria
BERN
Switz.
BUDAPEST
Hungary
Moldavian S.S.R.
Kishinev
Odessa
Sea of Azov
Kerch'
Crimean Peninsula
Simferopol'
Sevastopol'
Romania
BELGRADE
Yugoslavia
BUCHAREST
Danube
Black Sea
Corsica
Italy
ROME
Bulgaria
SOFIA
TIRANE
Alb.
Sardinia
ANKARA
Greece
Turkey
ATHENS
Sicily
TUNIS
Tunisia
Mediterranean Sea
200 Kilometers
200 Miles
Malta (U.K.)
NICOSIA
Cyprus (U.K.)
Boundary representation is not necessarily authoritative
~~Secret~~ NOFORN
724744 (B00997) 4-92

shutter ruined much of the photography of one of the flights. The third mission (2023), on the following day, included the Crimean Peinsula.[31]

The film from the first overflight (4 July) was flown to the United States immediately after the U-2 landed at Wiesbaden. Several members of the Photo Intelligence Division were on hand when the film was developed to check on the results. Also present was James Baker, who had accepted an offer by project officials to get a first-hand look at how the new A-2 lenses were working.[32]

The photos from July overflights were generally good, despite occasional problems caused by cloud cover. The huge amount of film taken by these missions provided more information about the Soviet Union's ability to track and intercept U-2s. Photointerpreters examining the films eventually discovered the tiny images of MiG-15s and MiG-17s beneath the U-2s in various pursuit and attack attitudes: climbing, flipping over, and falling toward Earth. It was even possible to determine their approximate altitudes. These photographs showed that the Soviet air defense system was able to track U-2s well enough to attempt interception, but they also provided proof that the fighter aircraft available to the Soviet Union in 1956 could not bring down a U-2 at operational altitude.[33]

One problem with early U-2 photography became apparent only after the first films were developed. If there was surface water on the runway at Wiesbaden when the U-2 took off, the camera windows became begrimed. Although the water dried during the flight, the oily scum it left behind degraded the photographic image. To combat this problem, AQUATONE ground crews took brooms and spent several hours before takeoff sweeping puddles of water from the runway to be used by the U-2. Kelly Johnson eventually designed a jettisonable cover for the camera windows, which was released at the same time as the pogos so that it could be recovered and reused.[34]

---

[31] Mission folders 2020 (9 July 1956), 2021 (9 July 1956), and 2023 (10 July 1956), OSA records, job 67-B-328, box 7 (~~TS~~ Codeword).

[32] Cunningham interview (~~TS~~ Codeword).

[33] Lundahl and Brugioni interview (~~TS~~ Codeword).

[34] Baker interview (~~S~~).

SOVIET PROTEST NOTE

The 4 and 5 July overflights brought a strong protest from the Soviet Union on 10 July in the form of a note handed to the US Embassy in Moscow. The note said that the overflights had been made by a "twin-engine medium bomber of the United States Air Force" and gave details of the routes flown by the first two missions. The note did not mention Moscow or Leningrad, however, because the Soviets had not been able to track these portions of the overflights. The Soviet note stated that the flights could only be evaluated as "intentional and conducted for the purposes of intelligence." As soon as the note arrived at the White House on the evening of 10 July 1956, Colonel Goodpaster called Bissell and told him to stop all U-2 overflights until further notice. The next morning Goodpaster met with Bissell to review the U-2 situation. Bissell said three additional flights had taken place since the missions mentioned in the Soviet note but added that no more were planned.[35]

Later Eisenhower told Goodpaster that he "didn't like a thing" about the Soviet note and was going to discuss the matter with Secretary of State Dulles. With the strong approval of President Eisenhower, Goodpaster informed DCI Dulles that "there is to be no mention of the existence of this project or of operations incident to it, outside the Executive Branch, and no mention within the Executive Branch to others than those who directly need to know of the operation, as distinguished from output deriving from it." [36]

During these initial overflights, the U-2 flew above 69,000 feet and could be seen only fleetingly by pilots of the Soviet interceptor aircraft. Thus, it appears that the Soviet claim that the intruder was a twin-engine bomber was probably based on the assumption that this was another overflight by a reconnaissance version of the twin-engine Canberra bomber, similar to the RAF overflight of Kapustin Yar in 1953. The US reply, sent to the Soviets on 19 July, truthfully denied that any US "military planes" had overflown the Soviet Union on the days in question. Meanwhile, on 16 July the Polish Ambassador to

[35] "Alleged Violations of Soviet Territory: Soviet Note of July 10, 1956 with U.S. Reply," *US Department of State Bulletin*, 30 July 1956, pp. 191-192; Andrew J. Goodpaster, Memorandum for the Record, 11 July 1956, WHOSS, Alpha, DDEL ~~(PS)~~.

[36] Andrew J. Goodpaster, Memorandum for the Record, 11 July 1956, WHOSS, Alpha, DDEL ~~(PS)~~.

the United States delivered an oral protest concerning overflights of Poland on 20 June and 2 July. This was followed by a protest note from the Czechoslovak Government on 21 July. No formal reply was sent to the two Soviet satellite states.[37]

The details of the flightpaths listed in the Soviet and Polish protests, along with the subsequent photographic evidence of Soviet interception attempts, made it clear that U-2s could not fly undetected over the Soviet Union or Eastern Europe and could even be tracked for extended periods of time. This news greatly disturbed President Eisenhower. In a meeting with Allen Dulles on 19 July 1956, the President recalled how he had been told that "not over a very minor percentage of these (flights) would be picked up." He went on to question "how far this should now be pushed, knowing that detection is not likely to be avoided." After discussing the possibility of basing U-2s in the Far East, President Eisenhower went on to say that he had "lost enthusiasm" for the U-2 activity. He noted that, if the United States were on the receiving end of a Soviet overflight operation, "the reaction would be drastic." The President was also concerned that the American public might learn of the overflights and be shocked that their country had violated international law. He stated, "Soviet protests were one thing, any loss of confidence by our own people would be quite another."[38]

The President's rapid disenchantment with the project was not lost on Richard Bissell. Fearing for the U-2 program's survival, he met with the Land committee in early August 1956 to urge them to help make the U-2 less vulnerable to radar pulses. His goal was to reduce the aircraft's radar cross section so that it would be less susceptible to detection. Edward Purcell had some ideas on this and suggested that he supervise a new project in the Boston area to explore them. At the direction of the Land committee, Bissell set in motion a project known as HTNAMABLE to establish a proprietary firm called the Scientific Engineering Institute (SEI) in Cambridge. Former Air Force Col. Richard S. Leghorn headed the SEI operation for the Agency when it began on 26 November 1956. SEI was staffed by several MIT scholars who conducted studies and experiments into

[37] "Alleged Violations of Soviet Territory: Soviet Note of July 10, 1956 with U.S. Reply," *US Department of State Bulletin*, 30 July 1956, pp. 191-192; *OSA History*, chap. 11, pp. 32-33 (~~TS~~ Codeword).

[38] Andrew J. Goodpaster, Memorandum for the Record, 19 July 1956, WHOSS, Alpha, DDEL (~~TS~~).

radar-absorbing materials and techniques proposed by Purcell. The effort, known as Project RAINBOW, got under way by the end of the year.[39]

THE END OF THE BOMBER GAP

During the three-week period of 20 June to 10 July 1956, U-2s had made eight overflights beyond the Iron Curtain, including five over the Soviet Union. PID's photointerpreters were busy until the end of August with their initial evaluation of the photography obtained by these flights. Their efforts were complicated by the division's move on 9 July from Que Building to the Steuart Building, but, when the photointerpreters were finished, they were able to write "finis" to the controversy over Soviet bomber strength.

Although the Air Force had claimed that the Soviet Union possessed almost 100 of the new Myasishchev-4 (Bison) heavy bombers, U-2 photography proved this assertion wrong. There were no Bison bombers at any of the nine long-range bomber bases photographed by the July missions. DCI Allen Dulles was particularly impressed by the photographs of the Soviet bomber bases, which in later years he called "million-dollar" photography. The actual value of the U-2 photos was probably even greater because, on the strength of their evidence, the White House was able to deny Air Force requests for additional B-52 bombers to "catch up" to the Soviets.[40]

Because of the need to protect the source of the information about Soviet bomber strength, the controversy surrounding this issue did not immediately die down. In November 1956, when the CIA began providing new Bison production figures based on U-2 photography without identifying the source, some members of Congress—unaware of the existence of the U-2—questioned the motivation behind the reduced estimates. They suggested that either the earlier estimates of Soviet bomber strength had been inflated to increase Air Force appropriations or the new estimates had been reduced by White House direction in order to hold down military expenditures.

[39] Records of Scientific Engineering Institute (Project HTNAMABLE), OSA records (~~TS~~ Codeword).

[40] *NPIC History*, vol. 1, p. 23 (~~S~~).

No one in the White House, the CIA, or the Air Force could reveal that U-2 photographs had actually provided the primary evidence for this change in the estimates.[41]

The need to keep the existence of the U-2 program secret caused problems even within the CIA itself. The Office of Security sharply restricted the number of persons who could be cleared for access to U-2 photography. The special clearance was granted on a "slot" basis, and only the person assigned to a particular position or "slot" could have the clearance. The U-2 photographs were kept in a secure room, and only those with special clearances were admitted to the room. In addition, the Office of Security considered U-2 information too sensitive to use in CIA publications. As a result, many analysts did not have access to information that would have greatly aided the production of intelligence estimates.[42]

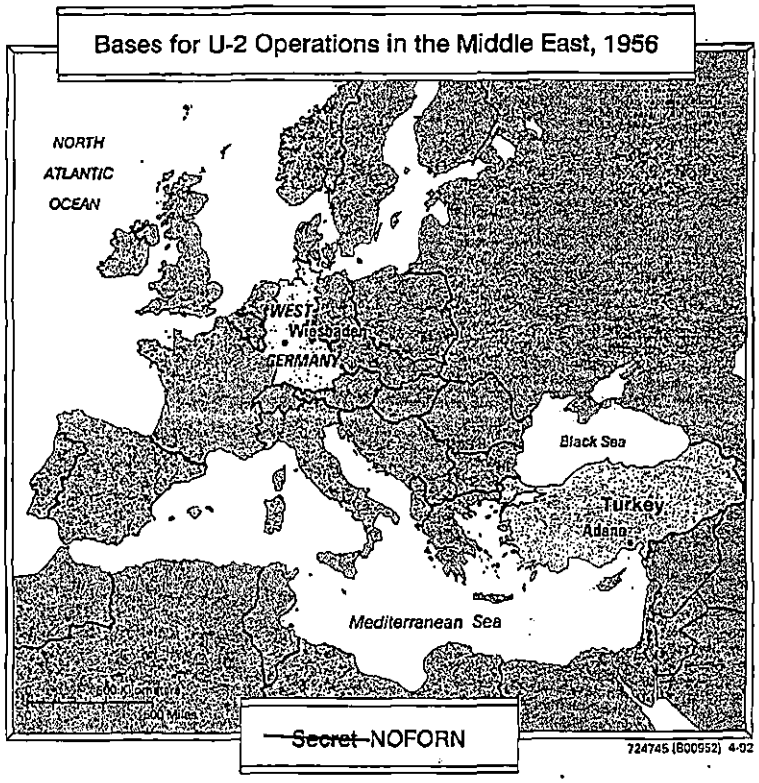
TACTICAL INTELLIGENCE FROM U-2s DURING THE SUEZ CRISIS

Although U-2s had ceased flying over the Soviet Bloc because of President Eisenhower's standdown order, they could still be used elsewhere in the world. The Middle East would be the next area for U-2 operations. On 26 July 1956, Egyptian President Gamal Abdel Nasser nationalized the Suez Canal Company in retaliation for the decision by the United States and the United Kingdom to withdraw financial support for the Aswan Dam project. Nasser's action provoked an international crisis that would have a permanent effect on the U-2 program.

Long before the Suez Crisis developed, the CIA had planned to deploy U-2s in Turkey for use in the Soviet overflight program. On 1 May 1956, US Charge d'Affaires Foy D. Kohler approached Turkish Prime Minister Adnan Menderes on this matter. He told the Prime Minister that the effort was a continuation of the GENETRIX program, during which balloons had been released from Turkey, and involved aircraft that could fly 10,000 feet higher than any Soviet plane. Menderes gave his approval immediately. At the time of the

[41] John Prados, *The Soviet Estimate: U.S. Intelligence Analysis and Russian Military Strength* (New York: Dial Press, 1982), pp. 45-47.

[42] Lundahl and Brugioni interview (TS Codeword).



Suez takeover, however, the second contingent of U-2 aircraft and pilots was still being trained in Nevada. This unit would not be ready for redeployment before the end of August and would not become established at Incirlik airbase near Adana, Turkey, until early September 1956. The Agency referred to the AQUATONE detachment at Adana as Detachment B, cryptonym KWCORK; the Air Force covername was Weather Reconnaissance Squadron Provisional 2; and the unit's unofficial name was Tuslog Detachment 10-10. By whatever name, the Adana detachment became the mainstay of U-2 activity for the next three and a half years.[43]

The fast-moving events of the Suez Crisis would not wait for Detachment B pilots to complete their training. With tension growing between Egypt and the Suez Canal Company's former owners, the United Kingdom and France, as well as between Egypt and Israel, US

[43] *OSA History*, chap. 11, pp. 9, 39-40; chap. 12, pp. 5, 12 (~~TS~~ Codeword).

military and foreign policy planners needed immediate information about developments in the eastern Mediterranean. Detachment A was, therefore, assigned the first Middle East overflights. On 29 August, U-2 missions 1104 and 1105 left Wiesbaden and overflew the eastern Mediterranean littoral, starting with Greece, then Egypt, Israel, Lebanon, Syria, and Turkey. Because these target areas were beyond the round trip range of the Wiesbaden-based U-2s, the planes landed at Adana for refueling. The next day, the same two planes, with different pilots, took off from Adana and overflew the same Middle East territory, this time including the Gaza Strip, before returning to Wiesbaden. The film contained evidence of large numbers of British troops on Malta and Cyprus and more new French-made aircraft in Israel than had previously been reported.[44]

As the situation around Suez grew more tense, the Eisenhower administration decided to release some of the U-2 photos to the British Government. On 7 September, James Reber, chairman of the Ad Hoc Requirements Committee, and Arthur Lundahl, chief of the Photo Intelligence Division, flew to London, taking with them photos of the eastern Mediterranean area, including the Suez Canal, taken on 30 August. These were the first and the only photos of the Middle East that the President authorized to be given to the British during the 1956 crisis.[45]

The Eisenhower administration viewed the developments in the eastern Mediterranean with great concern. To keep the President and Secretary of State abreast of developments in the area, Deputy Director for Intelligence Robert Amory established on 12 September a multiagency group known as the PARAMOUNT Committee to monitor the situation on a round-the-clock basis. The PARAMOUNT Committee worked inside PID headquarters in the Steuart Building. Composed of members from CIA, State, NSA, Army, Navy, and Air Force, this committee met daily—frequently several times a day—to produce reports based on information obtained from U-2 photography, communications and electronic intelligence, and agents. The photointerpreters working for the PARMOUNT committee also came from several agencies: the CIA, the Army, and the Navy.[46]

---

[44] Mission folders 1104 (29 August 1956) and 1105 (29 August 1956), OSA records, job 67-B-972, box 6 (~~TS~~ Codeword).

[45] Lundahl and Brugioni interview (~~TS~~ Codeword); *NPIC History*, vol. 1, pp. 56-58 (S).

[46] *NPIC History*, vol. 1, pp. 47-49, 54-56 (~~S~~).

The Suez Crisis was a major turning point in the use of the U-2 airplane. Before this crisis, the U-2 had been seen solely as a collector of strategic intelligence, with high-quality results considered more important than speed. U-2 film had, therefore, been returned to the manufacturer for optimum development and then interpreted in Washington using the most up-to-date devices. Now, because of the Middle East crisis, Project AQUATONE was expected to perform like a tactical reconnaissance unit, developing film immediately after landing for instant interpretation or "readout." Photo-Intelligence Division personnel assigned to Project HTAUTOMAT (U-2 film exploitation), therefore, had to arrange for forward processing of the U-2 film to avoid unacceptable delays in providing intelligence on tactical developments around Suez.

Withheld under statutory authority of the Central Intelligence Agency Act of 1949 (50 U.S.C., section 403g)

PID acted quickly to carry out its new assignment. Lundahl and Reber flew from the United Kingdom to US Air Force Europe headquarters in Wiesbaden on 12 September to make arrangements for processing and interpreting U-2 film in West Germany. They had been preceded by [redacted] chief of PID's Special Projects Branch. Following detailed discussions with Air Force photo-intelligence personnel, the CIA representatives arranged to use a portion of a nearby Air Force photo laboratory for developing U-2 film. With the assistance of [redacted] chief of the HTAUTOMAT photo laboratory, and Air Force personnel, [redacted] had the lab ready for processing on the following day, when the next U-2 mission returned from the Middle East. After quickly developing the film, [redacted] and his joint staff of CIA and armed forces personnel studied it for indications of British and French preparations for hostilities and sent their first report to Washington on 15 September.

Although the Air Force provided considerable assistance in establishing [redacted] photo laboratory, Air Force officials did not like the idea of CIA personnel controlling overseas photo processing and interpretation centers, which were normally under Air Force control. Further negotiations led to a CIA–Air Force agreement at the end of October, under which the Air Force would name the commanding officer for such installations and the CIA would designate the deputy, who was responsible for technical and intelligence matters.[47]

PID soon added two photointerpreters and a lab technician to the West German operation, which continued to develop and interpret U-2 photography of the Middle East throughout September and

[47] Ibid., pp. 49-52 (~~S~~).

October 1956. This unit's timely and accurate information enabled the PARAMOUNT Committee to predict the joint Israeli-British-French attack on Egypt three days before it took place.

On 11 September, Detachment A pilot Jacob Kratt overflew the French Mediterranean naval base at Toulon. He brought back imagery revealing that France was loading troopships at Toulon. During the rest of the month, Detachment A pilots flew another eight missions over the Middle East. By this time, the new Detachment B in Turkey was ready for operations, and it was better positioned to provide coverage of the Middle East. Detachment B began flying missions in September and soon became the primary detachment for Middle East overflights, conducting nine out of the 10 such missions flown in October.[48]

Detachment B's first U-2 flight, on 11 September 1956, made passes over Turkey, Cyprus, and Rhodes. The next flight, more than two weeks later, covered much the same ground but flew as far west as Malta, Sicily, and Crete. Both were "special" missions aimed at maintaining surveillance of the British and French fleets and forces as they prepared for the attack on Egypt. Meanwhile, Detachment A pilots flew four missions in the Western Mediterranean.

During this period, the PARAMOUNT Committee's photointerpreters developed the new science of "tent-ology"—counting the tents of British forces on the islands of Cyprus and Malta to determine the number of troops deployed. The photointerpreters used changes in the number of tents to determine that the British forces were beginning to move toward the beaches, where they were eventually lightered to offshore troopships.[49]

Noting the U-2 activity in the Middle East, President Eisenhower wrote in his diary on 15 October 1956: "Our high-flying reconnaissance planes have shown that Israel has obtained some 60 of the French Mystere pursuit planes, when there had been reported the transfer of only 24."[50] Other U-2 photographs revealed the presence

[48] *OSA History*, chap. 19, annex 120, "CIA U-2 Missions Flown, 1956-1968," pp. 1-2 (~~TS~~ Codeword).

[49] Lundahl and Brugioni interview (~~TS~~ Codeword).

[50] Dwight D. Eisenhower Diary, 15 October 1956, DDEL.

of British Canberra bombers at Akrotiri, Cyprus. The Anglo-French military buildup greatly irritated President Eisenhower, who considered these activities a violation of the 1950 Tripartite Declaration, in which the United States, the United Kingdom, and France had agreed to maintain the status quo in armaments and borders in the Middle East. To Arthur Lundahl he remarked, "It's a hell of a note when you have to G-2 your friends." [51]

U-2 photography continued to keep the President and other key officials well informed about the progress of the crisis. Flights over Israel and Rhodes on 21 October and Cyprus on 25 October revealed heavy military concentrations and an increase in the number of troop transports and air forces. On the basis of this information, Secretary of State John Foster Dulles told the President on 28 October that he believed an Israeli attack on Jordan was imminent, adding that he thought the British and French would take advantage of such an attack to occupy the Suez Canal.[52]

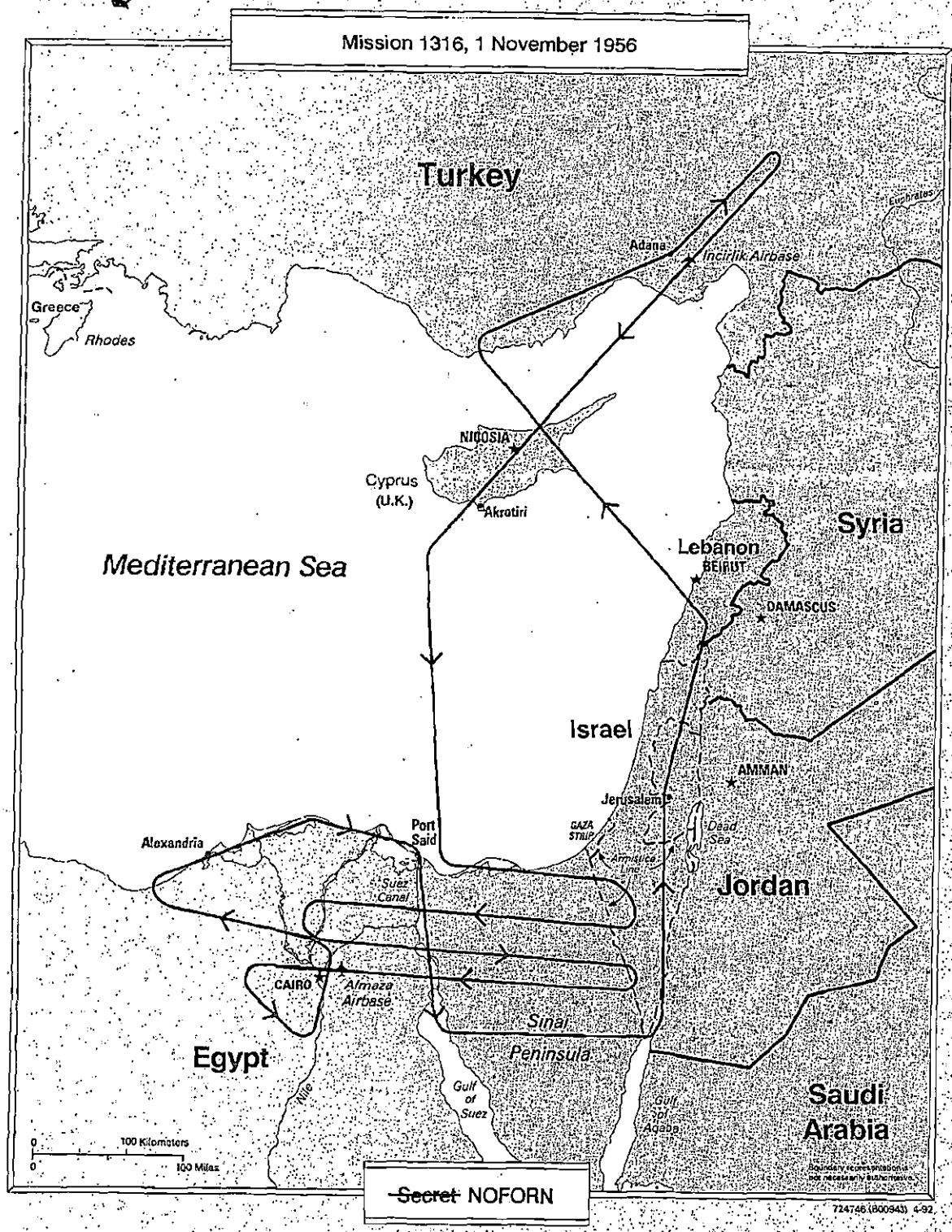
The 10-day Middle East war began on the afternoon of 29 October 1956 with Israeli paratroop drops in the Sinai peninsula, followed by mobile columns striking deep into Egyptian territory. The next day, 30 October, Francis Gary Powers conducted mission 1314. He overflew Lebanon, Jordan, Israel, and finally the Sinai, where he photographed black puffs of smoke from the fighting between Israel and Egypt. Adana-based U-2s were in the air for the next two days filming the Suez Canal area and neighboring countries.[53]

The United Kingdom and France entered the fray on the evening of 31 October with bombing raids against major Egyptian airfields. The Anglo-French bombing campaign continued for the next 48 hours. Early on the morning of 1 November, an Adana-based U-2, piloted by William Hall, took off to gather intelligence on the Anglo-French military activity. After photographing Cyprus, Hall flew south to the Sinai Desert, where he made several passes to obtain complete coverage of the Israeli-Egyptian fighting there. He then headed west to Cairo, passing directly over the main Egyptian airbase

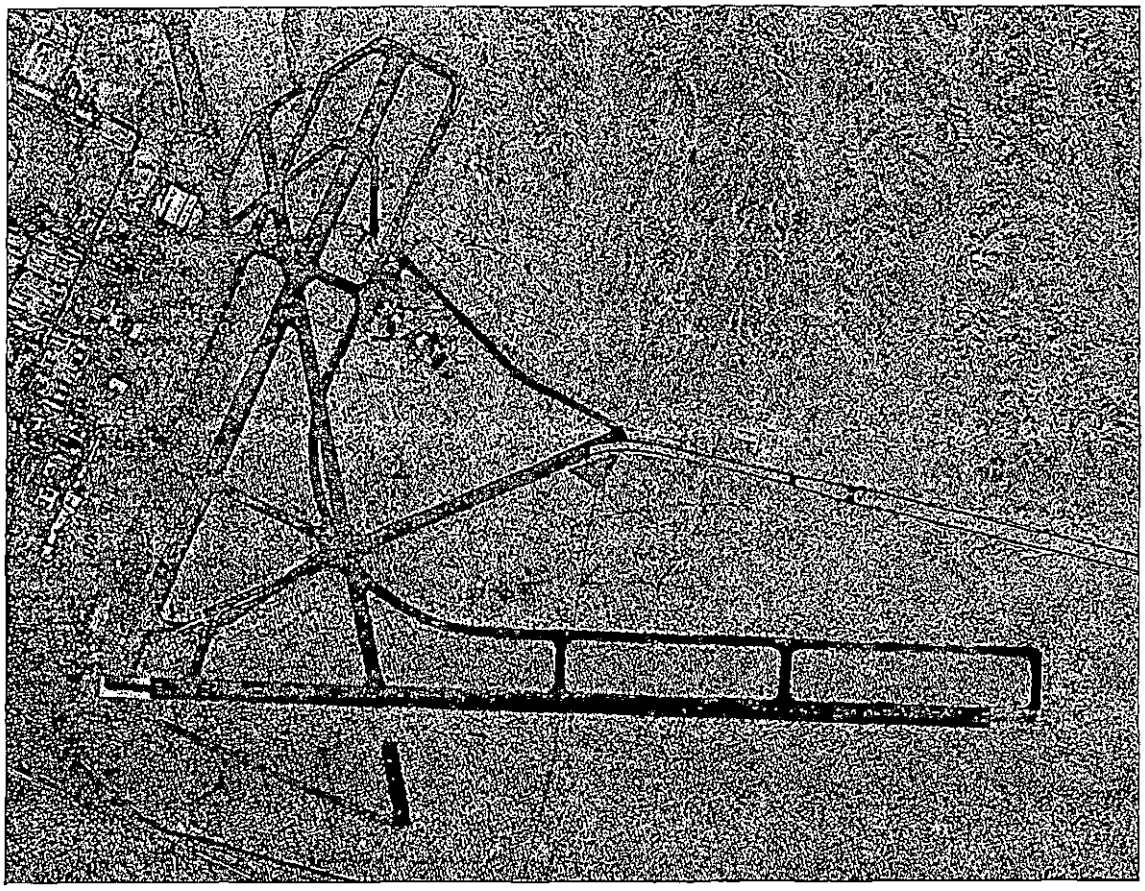
[51] Lundahl and Brugioni interview (~~TS~~ Codeword).

[52] Telephone calls, 28 October 1956, DDE Diary, DDEL.

[53] Powers, *Operation Overflight*, pp. 308-309; Mission folder 1314 (30 October 1956), OSA records, job 67-B-972, box 1 (~~TS~~ Codeword).



~~Secret~~ NOFORN



U-2 photography of Egyptian airbase at Almaza, 29 October 1956.

at Almaza, where he filmed neatly arranged rows of Egyptian military aircraft. Continuing past Cairo to film another airfield, Hall turned southeast and then north to fly along the Nile, again crossing directly over Almaza. The photography from this leg of the mission revealed the burning wreckage of the Egyptian aircraft. During the short period of time that had passed between Hall's two passes, a combined Anglo-French air armada had attacked the airbase. When shown the before and after photos of Almaza, President Eisenhower told Arthur Lundahl: "Ten-minute reconnaissance, now that's a goal to shoot for!"[54] Eisenhower was pleased with the aerial photography but

[54] Lundahl and Brugioni interview (~~TS~~ Codeword); Beschloss (*Mayday*, p. 138) mistakenly identifies this quote as coming from the British, but they did not receive copies of these photos.

angered by what it depicted: an Anglo-French attack on Egypt. He quickly called for a cease-fire and denied the United Kingdom any further U-2 photographs of the Middle East.

The 1 November mission over Cyprus and Egypt also photographed Anglo-French preparations to invade Egypt. President Eisenhower was informed of this impending invasion on Sunday, 4 November. On the following day, British and French paratroopers dropped near Port Said at the north end of the Suez Canal. This action prompted Soviet Premier Nikolai Bulganin to send messages to France, Britain, and Israel warning that the Soviet Union was ready to use force to crush the aggressors.[55]

Early on the morning of election day, 6 November, the Anglo-French invasion armada arrived at Port Said and began landing troops. Back in Washington President Eisenhower met with Allen Dulles to discuss the deepening international crisis. Worried that the Soviet Union might be poised to intervene in the war, the President ordered Dulles to have the Adana-based U-2s fly over Syria to see whether the Soviets were moving planes to Syrian airbases in preparation for a strike against the forces attacking Egypt. The answer to Eisenhower's question came much sooner than expected because on the previous day a U-2 had already overflown Syria before making a run across northern Egypt. The film from this flight had reached Wiesbaden for processing and readout during the night. The results were in the hands of the PARAMOUNT Committee by midmorning on 6 November, while the President was motoring to Gettysburg to cast his ballot. By the time the President returned to the White House by helicopter at noon, Colonel Goodpaster was waiting for him with an answer: there were no Soviet aircraft in Syria. Because of the President's concern about possible Soviet moves, Syria was the target of 14 additional U-2 flights between 7 November and 18 December 1956.[56]

The increasing reliance on Adana-based U-2s for coverage of the Middle East during the Suez Crisis made it difficult for the photointerpreters to supply timely information. When Detachment B aircraft returned to their base at Adana, there were no film-processing

[55] Donald Neff, *Warriors at Suez: Eisenhower Takes America into the Middle East* (New York: Simon and Schuster, 1981), p. 403.

[56] Memorandum of Conference, 6 November 1956, Eisenhower Diary, Whitman File, DDEL (U); *OSA History*, chap. 19, annex 120, p. 3 (~~TS~~ Codeword).

facilities available, and the film had to be flown to Wiesbaden, adding a 10- to 15-hour delay. During the gradual buildup of the crisis, this delay had been tolerated, but, once actual hostilities broke out, US decisionmakers needed a more rapid response. On 29 October, Richard Bissell ordered Lundahl to establish a film-processing facility at Adana. Two PID employees went to Adana on 13 November to set up the facility, and two photointerpreters moved from Wiesbaden to Adana to help in the effort. Forward processing was, however, hampered by the location of the Adana facility on a flat, arid plain in southern Turkey, 35 miles from the Mediterranean at the very end of a long supply line.

The PID team obtained and outfitted a trailer for film processing, but many problems had to be overcome. The first major problem was obtaining enough clean water. Detachment B personnel, therefore, purchased large amounts of borax locally for use in purifying water. In fact, they bought so much borax on the local market that one of them was arrested by the Turkish police, who believed he was using the chemical to make drugs. It was also difficult to obtain a constant source of developers and fixers for processing the U-2 film, since the large Air Force supply facility at Wheelus AFB in Libya refused to provide the needed photographic chemicals. When PID personnel accompanied processed film from Turkey to the United States, they returned to Turkey sitting atop cartons of chemicals for the next day's processing. At first, film was developed in improvised tanks using flimsy wooden spools and hand-turned cranks to move the film through the solutions. Later, the Adana facility moved from its trailer to a building and received more up-to-date processing equipment. As was the case with the photo lab in Germany, the Adana lab's personnel came from the Agency and the armed forces.[57]

The need to produce very timely intelligence diminished after the British and French agreed to a cease-fire on 7 November 1956. By the end of the month, foreign troops began evacuating Egyptian territory, and the pressure on the Adana photointerpretation unit eased. The facility remained in existence, however, and was used twice in December 1956 and 11 times in the first half of 1957. It was then placed in caretaker status, for emergency use only.

[57] *NPIC History*, vol. 1, pp. 53-54 (~~S~~); Lundahl and Brugioni interview (~~TS~~ Codeword).

RENEWED OVERFLIGHTS OF THE SOVIET UNION

Throughout the fall of 1956, U-2s provided valuable coverage of the Middle East crisis, but they were not conducting their original mission of strategic reconnaissance of the Soviet Union. President Eisenhower had halted all such overflights by his order of 10 July, and, in the months that followed, he remained unconvinced by CIA arguments in favor of a resumption of overflights. On 17 September 1956, DDCI Cabell and Richard Bissell went to the White House to ask President Eisenhower to authorize more flights over the Soviet Union. Adm. Arthur W. Radford, Chairman of the Joint Chiefs of Staff, also attended the meeting. Bissell and Defense Department representatives reviewed the valuable intelligence from the July U-2 flights, and Bissell then informed Eisenhower that many important intelligence requirements remained unfilled. To fill these requirements, Bissell noted, would require photography of approximately 15 separate areas of the Soviet Union. Pleading for the authority to resume overflights, Bissell stressed that conditions for photography were becoming less favorable as the days grew shorter. While the U-2 was then still safe from interception, he added, it might not be in the future.[58]

President Eisenhower acknowledged the value of the U-2 but emphasized that the international political aspects of overflights remained his overriding concern. He said he would talk further with John Foster Dulles about the matter, noting that the Secretary of State had at first seemed to belittle the political risk but had later found it increasingly worrisome.

A little more than two weeks later, on 3 October, when the President again met with Bissell, Cabell, and Radford, John Foster Dulles was also present. In opening the meeting, Eisenhower said he had become discouraged regarding Project AQUATONE. Although he had been assured that "there would be a good chance of not being discovered on most, if not all, operations, just the opposite had proved true." The President observed that arguments in favor of resuming U-2 operations did not take world opinion into consideration. He added that great efforts had been made for many years "to create an opinion in the world that we are not truculent and do not want war," and, if knowledge of the U-2 overflights got out, world opinion would view them as "provocative and unjustified."[59]

---

[58] Andrew J. Goodpaster, Memorandum of Conference, 17 September 1956, WHOSS, Alpha, DDEL (~~TS~~).

[59] Andrew J. Goodpaster, Memorandum for the Record, 3 October 1956, WHOSS, Alpha, DDEL (~~TS~~).

Secretary of State Dulles said that, although he essentially agreed with the President's comments, he thought that "really important results" might be obtained by a seven to 10-day operation. He, nevertheless, questioned the long-term value of the results. DDCI Cabell replied that U-2 photographs would be useful much longer than the Secretary of State had implied because they would establish a reference bank of geographic and manmade features. Siding with Cabell, Admiral Radford pointed out the need for more intelligence to make estimates better.

President Eisenhower was not convinced by these arguments. Although willing to consider extensions of the radar-seeking ferret flights he had authorized along the Soviet borders, he remained opposed to penetration flights over the Soviet Union.

Events in Eastern Europe in the fall of 1956 helped to change the President's mind. In October the Soviet Union backed away from a confrontation with nationalist Communist leaders in Poland only to find itself facing a similar situation in Hungary, where mass demonstrations led to the formation of a new government under Imre Nagy on 23 October 1956. Soviet troops and tanks temporarily withdrew from Budapest while awaiting reinforcements. By early November, however, the Kremlin leadership decided that events in Hungary were getting out of hand—particularly when Premier Nagy announced his nation's withdrawal from the Warsaw Pact—and ordered Soviet troops to suppress the Hungarian uprising. Although President Eisenhower deplored the Soviet intervention, he turned down CIA requests for permission to airdrop arms and supplies to the Hungarian rebels. In fact, the President forbid all overflights of that nation, including those by U-2 aircraft, and none was made.[60]

Although President Eisenhower had not been willing to allow overflights during the Hungarian crisis, the Soviet Union's actions in Hungary convinced him to authorize renewed overflights of the Soviet Bloc, a decision that was made easier by his reelection by a large margin in early November. Initially, however, the President only authorized overflights of Eastern Europe and Soviet border regions, not the deep penetration overflights that had been requested by CIA. At a 15 November 1956 meeting with Acting Secretary of State Herbert Hoover, Jr. (John Foster Dulles was recovering from cancer

[60] Cunningham interview (~~PS~~ Codeword).

surgery), JCS Chairman Adm. Arthur Radford, DCI Allen Dulles, and Richard Bissell, Eisenhower explained why he refused to allow overflights of the Soviet Union: "Everyone in the world says that, in the last six weeks; the United States has gained a place it hasn't held since World War II. To make trips now would cost more than we would gain in form of solid information." Hoover agreed and noted, "If we lost a plane at this stage, it would be almost catastrophic." Torn between his desire to maintain a "correct and moral" position and his wish to know what the Soviet Union was up to, the President finally authorized several overflights of Eastern Europe and the Soviet border, "but not the deep one," adding that the aircraft should "stay as close to the border as possible."[61]

The first of these flights, mission 4016 on 20 November 1956, was the first overflight of Soviet territory since 10 July. This mission left Adana and flew east over Iran, then reversed and flew west along the Soviet-Iranian border to Soviet Armenia, where it crossed into the Soviet Union and photographed Yerevan. An electrical malfunction then forced the pilot, Francis Gary Powers, to return to Adana. Soviet interceptor aircraft made several unsuccessful attempts to reach this U-2, and the Soviet Government sent a secret protest note to Washington.[62]

On 10 December, Bulgaria was the target of two U-2 missions, one (4018) from Detachment B at Adana and another (2029) from Detachment A at Giebelstadt. Bulgarian fighter aircraft made 10 different attempts to intercept the first mission, but the flight proceeded without difficulty.[63]

The second flight came close to crashing but not through the efforts of interceptors. The pilot of mission 2029 was Carmine Vito, who had flown the first U-2 mission over Moscow on 5 July. He was known to his colleagues as the Lemon-Drop Kid because he always carried these hard candies in the knee pocket of his flight suit. Despite

[61] Andrew J. Goodpaster, Memorandum of Conference with the President, November 15, 1956, WHOSS, Alpha, DDEL ~~(TS)~~; Ambrose, *Eisenhower: The President*, p. 374.

[62] Mission folder 4016, 20 November 1956, OSA records, job 67-B-972, box 7 ~~(TS~~ Codeword).

[63] Mission folders 2029 (10 December 1956) and 4018 (10 December 1956), OSA records, job 67-B-328, box 7 and job 67-B-972, box 7 ~~(TS~~ Codeword).

warnings to all pilots about the danger of opening the helmet faceplate at high altitudes, several pilots were known to do so. Some ate candy bars; Vito favored lemon drops. On the morning of 10 December, while Vito was undergoing prebreathing, the Air Force enlisted man who oversaw his preflight regimen placed an L-pill in the righthand knee pocket of Vito's flight suit, unaware that this pocket also contained Vito's supply of lemon drops. After he took off, Vito began indulging in his habit of sucking lemon drops. About midway into the mission, he opened his faceplate and popped into his mouth what he thought was another lemon drop. After closing the faceplate, he began sucking on the object and thought it strange that it had no flavor and was much smoother than the previous lemon drops. Although tempted to bite down, Vito decided instead to reopen his faceplate and see what it was he had in his mouth. Spitting the object into his hand, he saw that he had been sucking on the L-pill with its lethal contents of potassium cyanide. Just a thin layer of glass had stood between him and death. The loss of his aircraft over Bulgaria would have exposed the U-2 program to worldwide publicity and would probably have resulted in an early end to overflights.[64]

Detachment A's security officer overheard Vito relating the L-pill story to a fellow pilot several days later and promptly reported the conversation to headquarters. When details of Vito's close call reached Washington, James Cunningham immediately ordered L-pills placed in boxes so that there would be no chance of mistaking them for anything else. The L-pill continued to be available for another three years. Then in January 1960, the commander of Detachment B, Col. William Shelton, raised an important question that had never been considered: what would happen if an L-pill with its volatile contents accidentally broke inside the cockpit of a U-2? Realizing that such an accident would result in the death of the pilot, James Cunningham ordered the destruction of all L-pills and then turned to the Technical Services Division (TSD) for a better idea. By this time the state of the art in lethal devices was a needle poisoned with algal, an extremely deadly shellfish toxin. The needle was hidden in a tiny hole in a silver dollar supplied by Cunningham. Only one poison-needle coin was made because Cunningham decided that, if any pilot had to use it because of capture, there would probably not be any more overflights.[65]

---

[64] Cunningham interview (~~TS~~ Codeword); Carmine Vito, interview by Donald E. Welzenbach, Washington, DC, 7 May 1986 (~~S~~).

[65] Cable from Detachment B to Development Projects Division (DPD), 4 January 1960; cable from DPD to Detachment B, 7 January 1960, OSA records, job 67-B-972, box 18, "Operation KNIFE EDGE" (~~TS~~ Codeword).

Although the U-2 overflights of Eastern Europe in late 1956 caused renewed Soviet protests, the sharpest protest came on 15 December 1956, after three specially modified USAF RB-57D bombers photographed the city of Vladivostok in a high-speed dash over the Far Eastern coast of the Soviet Union (as part of the Air Force's Operation BLACK KNIGHT). President Eisenhower had approved the mission after being told by the Air Force that the high-speed RB-57Ds would probably not be detected.[66]

Reacting strongly to the Soviet protest, the President told Secretary of State Dulles on 18 December that he was going to "order complete stoppage of this entire business." As for a reply to the Soviet protest, Dulles said, "I think we will have to admit this was done and say we are sorry. We cannot deny it." Dulles noted that "our relations with Russia are pretty tense at the moment." Eisenhower agreed, noting that this was no time to be provocative. He then instructed Colonel Goodpaster to call Secretary of Defense Wilson, Joint Chiefs of Staff Chairman Radford, and DCI Dulles to order: "Effective immediately, there are to be no flights by US reconnaissance aircraft over *Iron Curtain* countries."[67]

Flights along the borders of Iron Curtain countries continued, however, and, on 22 December 1956, Detachment B flew the first mission (4019) by a U-2 equipped for electronic intercept. The electronic-detection equipment known as the System-V unit (see appendix C) was installed in the bay normally used by the main camera, and the plane flew along the Soviet border from the Black Sea to the Caspian Sea and on to Afghanistan. The System-V unit worked well.[68]

Early in 1957, a mission along the Soviet border accidentally turned into an overflight. On 18 March 1957, a U-2 collecting electronic intelligence along the Soviet southern border entered Soviet

---

[66] Goodpaster interview (~~S~~).

[67] Telephone calls 18 December 1956, DDE Diary, DDEL, (U); Andrew J. Goodpaster, Memorandum for the Record, 18 December 1956, WHOSS, Alpha, DDEL (~~TS~~; downgraded to ~~S~~); the Soviet protest note of 15 December 1956 and the U.S. reply of 11 January 1957 are contained in "Alleged Overflight of Soviet Area by American Planes," *US Department of State Bulletin*, vol. 36, 28 January 1957, p. 135. Although Dulles's initial inclination had been to offer an apology, the U.S. reply stated that the "only authorized United States Air Force flights in the general area of the Sea of Japan were normal training activities."

[68] Mission folder 4019 (22 December 1956), OSA records, job 67-B-972, box 7 (~~TS~~ Codeword).

airspace because of compass error compounded by a slight error in the pilot's dead reckoning. Because of heavy cloud cover, the pilot, James W. Cherbonneaux, did not realize he was over the Soviet Union until he saw Soviet fighters attempting to intercept him. These attempts at interception once again demonstrated the Soviets' ability to track the U-2 and their inability to harm it.[69]

At this point in early 1957, the U-2 program was in limbo. Although the President would not allow U-2s to fly their primary mission of reconnaissance of the Soviet Union, he did not cancel the program and continued to authorize flights along Soviet borders. The CIA's overhead reconnaissance program also faced a renewed bid by the Air Force, which now had its own growing U-2 fleet, to gain control of the overflight program in the spring of 1957. The uncertainty surrounding the future of the project made planning and budgeting extremely difficult. In April 1957, Richard Bissell asked the DCI and DDCI to push for a decision on whether the U-2 program was to continue in civilian hands and what its scope was to be. In briefing papers prepared for the DCI, Bissell argued for maintaining a nonmilitary overflight capability, which could "maintain greater security, employ deeper cover, use civilian pilots, keep the aircraft outside military control, and, therefore, make possible more plausible denial of US military responsibility in the face of any Soviet charges." In urging the resumption of overflights, Bissell stated that four U-2 missions over border regions of the Soviet Union or Eastern Europe had been detected by the Soviets without causing any diplomatic protest. He also noted that the President's Board of Consultants on Foreign Intelligence Activities had unanimously recommended the resumption of overflights.[70]

All of these issues were discussed on 6 May 1957, when President Eisenhower met with Deputy Secretary of Defense Donald Quarles, Air Force Chief of Staff Nathan Twining, Acting Secretary of State Christian Herter, and three CIA officals—DCI Dulles, DDCI Cabell, and Richard Bissell. The President expressed concern about the impact of overflights on US-Soviet relations and about possible Soviet responses such as closing off access to Berlin. Although

[69] Information supplied by James W. Cherbonneaux to Donald E. Welzenbach (S); Mission folder 4020 (18 March 1957), OSA records, job 67-B-972, box 7 (~~TS~~ Codeword).

[70] *OSA History*, chap. 4, pp. 15-16; annex 22 (~~TS~~ Codeword).

remaining opposed to flights over most of the Soviet Union, Eisenhower finally agreed to permit some flights over peripheral areas such as Kamchatka Peninsula and Lake Baikal, as well as the Soviet Union's atomic testing area at Semipalatinsk. Such overflights could be staged from Pakistan if the Pakistani Government consented. The President rejected the Air Force's request to take over the U-2 program, stating that he preferred to have the aircraft manned by civilians "during operations of this kind."[71]

The President had once again agreed to allow overflights of the Soviet Union, although only over certain areas, because the need to learn more about the capabilities and intentions of the Soviet Union was too compelling. In particular, the President and top administration officials wanted to gather more data on the Soviet Union's missile program, a subject for which considerable Soviet boasting—but no hard data—was available.

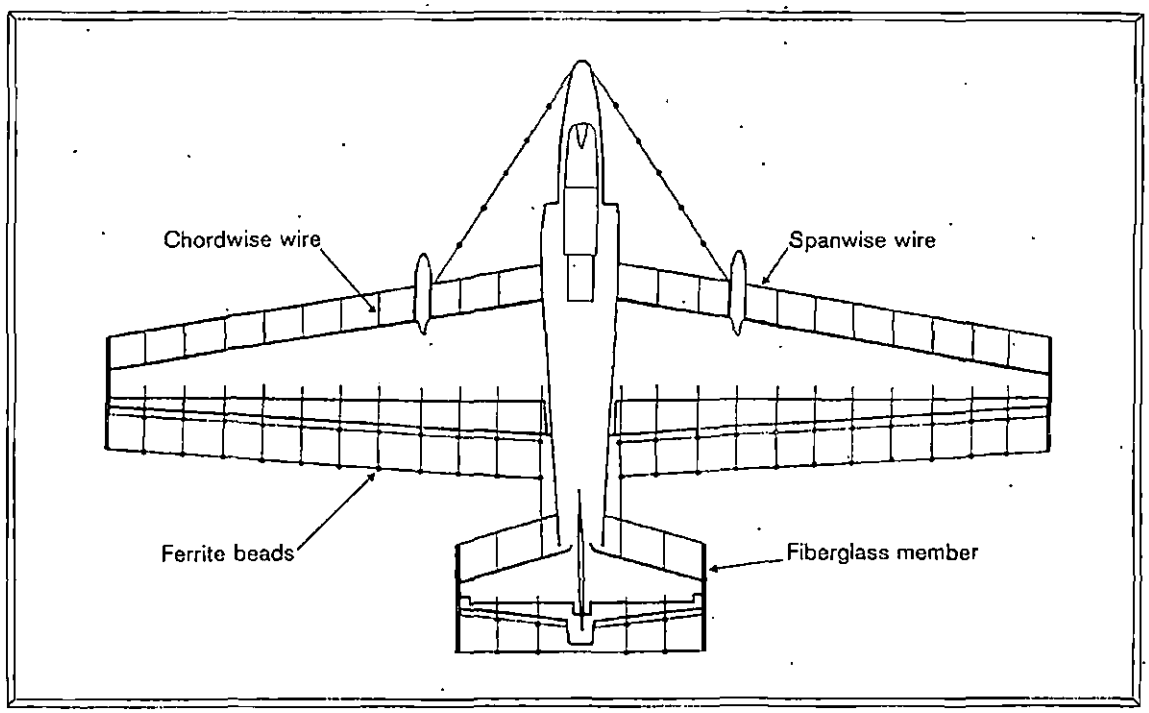
Even after he had authorized the resumption of overflights, President Eisenhower maintained tight control over the program. He personally authorized each overflight, which meant that Richard Bissell would bring maps to the White House with the proposed routes marked on them for the President to examine. More than once, according to Bissell, Eisenhower spread the map out on his Oval Office desk for detailed study, usually with his son John (an Army officer serving as a White House aide) and Colonel Goodpaster looking over his shoulder. On occasion, the President would pick up a pencil and eliminate a flight leg or make some other correction to the flight plan.[72]

RADAR-DECEPTIVE "DIRTY BIRDS"

One additional reason why President Eisenhower had again authorized overflights of the Soviet Union was renewed CIA promises that Soviet detection or tracking of the U-2 was unlikely. At the 6 May 1957 meeting with the President, Richard Bissell reported on the progress that had been made in developing radar camouflage and

[71] Andrew J. Goodpaster, Memorandum of Conference with the President, 6 May 1957 (TS); "Record of Action—Meeting of May 6, 1957," WHOSS, Alpha, DDEL ~~(TS)~~.

[72] Bissell interview by Welzenbach (S); Beschloss, *Mayday*, p. 140.



"Trapeze" antiradar attachments to the U-2

absorption devices for the U-2. Once these devices were installed on the operational U-2s, he explained, the "majority of incidents would be undetected." [73]

Work on methods of reducing the U-2's vulnerability to radar detection had begun in the fall of 1956 as the result of President Eisenhower's disenchantment with the overflight program following Soviet detection and tracking of the first series of U-2 missions. The CIA proprietary firm Scientific Engineering Institute was conducting this research under a project codenamed RAINBOW. SEI Chief Engineer Dr. Franklin A. Rodgers, formerly of MIT, converted the theories of Harvard physicist Edward Purcell into systems that could be used on aircraft. SEI's radar-deception system consisted of a series of attachments to the U-2. First bamboo poles and later fiberglass rods were attached to the wings, where they would not interfere with the control surfaces. At the ends of these poles, completely circling

[73] Andrew J. Goodpaster, Memorandum of Conference with the President, 6 May 1957 ~~(TS)~~; "Record of Action—Meeting of May 6, 1957," WHOSS, Alpha, DDEL ~~(TS)~~.

the aircraft, was a small-gauge wire with precisely spaced ferrite beads. The wire and beads were supposed to capture incoming 70-MHz radar pulses and either trap them in the loop or weaken them so much that they would not register as a valid radar return. This configuration was called the trapeze and was not very successful.

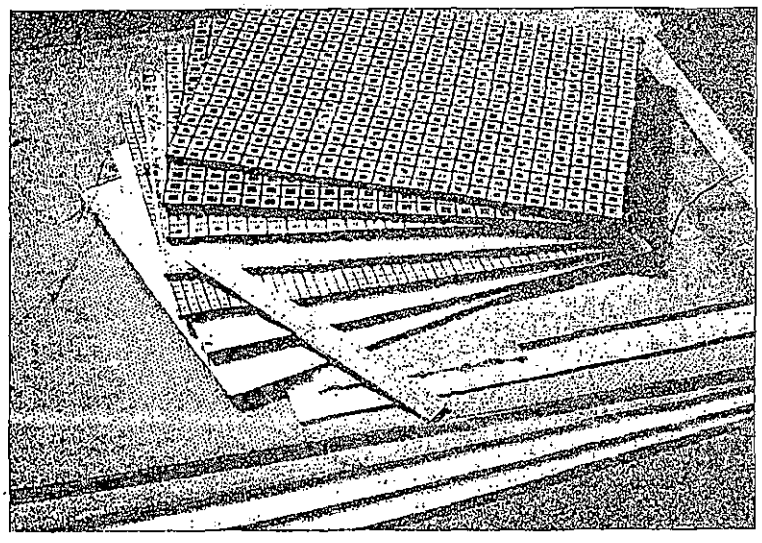
A second approach, tested in early 1958, involved the use of plastic material containing a printed circuit designed to absorb radar pulses in the 65- to 85-MHz range. Nicknamed "wallpaper," this material was glued to parts of the U-2's fuselage, nose, and tail. Although the "trapeze" and "wallpaper" systems provided protection against some Soviet radars, the systems proved ineffective against radars operating below 65-MHz or above 85-MHz. Furthermore, both of these additions degraded the U-2's performance. The weight and drag of "trapeze" reduced the aircraft's operating ceiling by 1,500 feet, and "wallpaper" sometimes caused engines to overheat.[74]

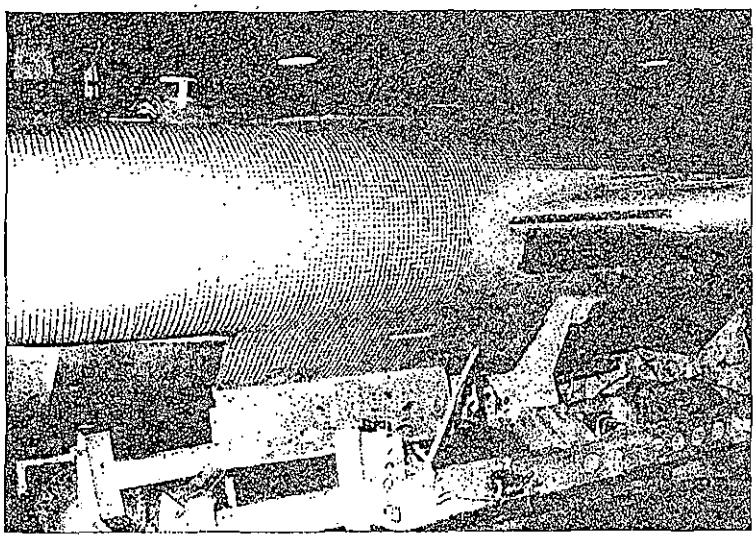
SEI's research results were tested by another firm known as Edgerton, Germeshausen & Grier (EG&G), which was also composed of MIT faculty members. Under an Air Force contract to evaluate radars, EG&G operated a small testing facility at Indian Springs, Nevada, not far from Area 51. Although Kelly Johnson had been closely involved with the radar deception project since its early days, he cooperated reluctantly because he disliked adding attachments that made his aircraft less airworthy. (Johnson's dislike of the antiradar attachments was reflected in the unofficial nickname for aircraft that had been so modified—"dirty birds.") After Lockheed mechanics had mounted the various RAINBOW devices on the prototype U-2, a Lockheed test pilot would fly the plane over EG&G's Indian Springs installation. This was little more than a series of radar sets and a trailer containing instrumentation. EG&G technicians could thus record and evaluate the U-2's radar returns as it traversed a specified course over their facility.[75]

This method of testing radar-deceptive modifications proved both time consuming and dangerous. During a test flight on 2 April 1957, the "wallpaper" modification acted as insulation around the

[74] Records of the Scientific Engineering Institute (Project HTNAMABLE), OSA records (~~TS~~ Codeword).

[75] References to EG&G programs for the U-2 are contained in the later Convair contracts for Projects FISH and KINGFISH, OSA records, job 67-B-415, box 1 (~~TS~~ Codeword).





"Wallpaper"

engine of the U-2 known as article 341, causing it to overheat and flameout. Unable to restart the power plant, Lockheed test pilot Robert Sieker bailed out but was struck and killed in midair by the U-2's tailplane. The aircraft crashed in an area of Nevada so remote that Area 51 search teams needed four days to locate the wreckage. The extensive search attracted the attention of the press, and a

Wreckage of Article 341, 2 April 1957

12 April 1957 article in the *Chicago Daily Tribune* was headlined, "Secrecy Veils High-Altitude Research Jet; Lockheed U-2 Called *Super Snooper*."[76]

Because of its large wingspan, an out-of-control U-2 tended to enter a classical flat spin before ground contact. This slowed descent and actually lessened the impact. If there was no fire after impact, the remains of crashed U-2s were often salvageable, as was the case with the wreckage of article 341. Kelly Johnson's crew at the Skunk Works used the wreckage, along with spares and salvaged parts of other crashed U-2s, to produce another flyable airframe for about $185,000.[77] The U-2's ability to survive a crash in fairly good condition should have been noted by the Development Projects Staff for consideration in its contingency plans for a loss over hostile territory because the equipment on board the aircraft could easily compromise the weather research cover story.

The loss of one of Lockheed's best test pilots, as well as the prototype "dirty bird" U-2, led Kelly Johnson to suggest that Lockheed install a large boom at the Indian Springs radar test facility. Using the

[76] Accident folder, crash of 2 April 1957, OSA records ~~(S)~~.

[77] Lockheed contracts, OSA Records ~~(S)~~.

boom, which could lift entire airframes 50 feet in the air, technicians could change the airframe's attitude and run radar tests almost continuously without having to fuel and fly the plane.[78]

By the summer of 1957, testing of the radar-deception system was complete, and in July the first "dirty bird" (DB) arrived at Detachment B. The first operational use of this aircraft occurred on 21 July 1957 in mission 4030 over Iran, Iraq, and Syria. On 31 July, the same aircraft made a run over the Black Sea. There were a total of nine DB missions over the USSR. The antiradar system did not prove very effective, and its use was curtailed in May 1958.[79]

THE NEW DETACHMENT C

On 8 June 1957, a U-2 took off from Eielson Air Force Base in Alaska to conduct the first intentional overflight of the Soviet Union since December 1956. This mission broke new ground in two respects: it was the first overflight conducted from American soil and the first by the new Detachment C.

Detachment C (known officially as Weather Reconnaissance Squadron, Provisional-3) was composed of the third group of pilots to complete their training in Nevada. In the autumn of 1956, this third detachment needed a new base because Area 51 was about to become the training site for a large number of Air Force pilots who would fly the 29 U-2s purchased by the Air Force. The Agency decided that the best location for Detachment C would be the Far East and began looking for bases there.

Even without the arrival of the Air Force pilots, Detachment C could not have stayed in Nevada much longer. In June 1957, the entire facility had to be evacuated because the Atomic Energy Commission was about to conduct a series of nuclear tests whose fallout was expected to contaminate the Groom Lake facility. All remaining CIA personnel, materiel, and aircraft were transferred to Edwards AFB, California, and became known as Detachment G.

[78] Ibid (~~S~~).

[79] Cunningham interview (~~TS~~ Codeword).

The search for a new home for Detachment C led the Agency to ask the Air Force in the autumn of 1956 for permission to locate the detachment at Yokota AFB, Japan. Because Yokota was already the base for one covert project (the very secret Air Force Project BLACK KNIGHT using RB-57s), Air Force Chief of Staff Twining did not wish to locate another one there and denied the request. The Agency then turned to the Navy, which granted permission for Detachment C to use the Naval Air Station at Atsugi, Japan. The Japanese Government received no notification of the proposed deployment because at that time it had no control over activities involving US military bases in Japan. Deployment of Detachment C began in early 1957 but was complicated by a recent decision to permit the families of Project AQUATONE employees to accompany them on overseas tours. As a result, program managers had to find housing facilities on the base or in nearby communities, not an easy task in crowded Japan.[80]

Detachment C began conducting missions in June 1957 after several aircraft and pilots flew to Eielson Air Force Base near Fairbanks, Alaska. Air Force radar order-of-battle reports and NSA studies had revealed that the radar network in the Soviet Far East, with antiquated radar sets and personnel of a lower caliber than those in the western Soviet Union, was relatively ineffective. To take advantage of these weaknesses, Detachment C staged three missions from Alaska into the Soviet Far East. The first, on 7/8 June (the aircraft crossed the international date line during the flight), was unable to photograph its target, the ICBM impact area near Klyuchi on the Kamchatka Peninsula, because of bad weather and, therefore, never entered Soviet airspace. A second attempt to photograph Klyuchi on 19/20 June was marred by a camera malfunction that ruined every third frame of photography. This flight was tracked by Soviet radars, but there was no attempt at interception. After a pause of almost three months during which Detachment C received a dirty-bird U-2, the detachment's third mission over Klyuchi on 15/16 September 1957 achieved excellent results. The radar-deception devices proved ineffective, however, as the U-2 was tracked by Soviet radar and trailed by five fighters.[81]

---

[80] *OSA History*, chap. 15, pp. 2, 16-19; chap. 16, p. 1 (~~TS~~ Codeword).

[81] Mission folders 6002 (8 June 1957), 6005 (20 June 1957), and 6008 (16 September 1957), OSA records, job 67-B-328, box 7 (~~TS~~ Codeword).

Withheld under statutory authority of the Central Intelligence Agency Act of 1949 (50 U.S.C., section 403g)

25X1, E.O.13526

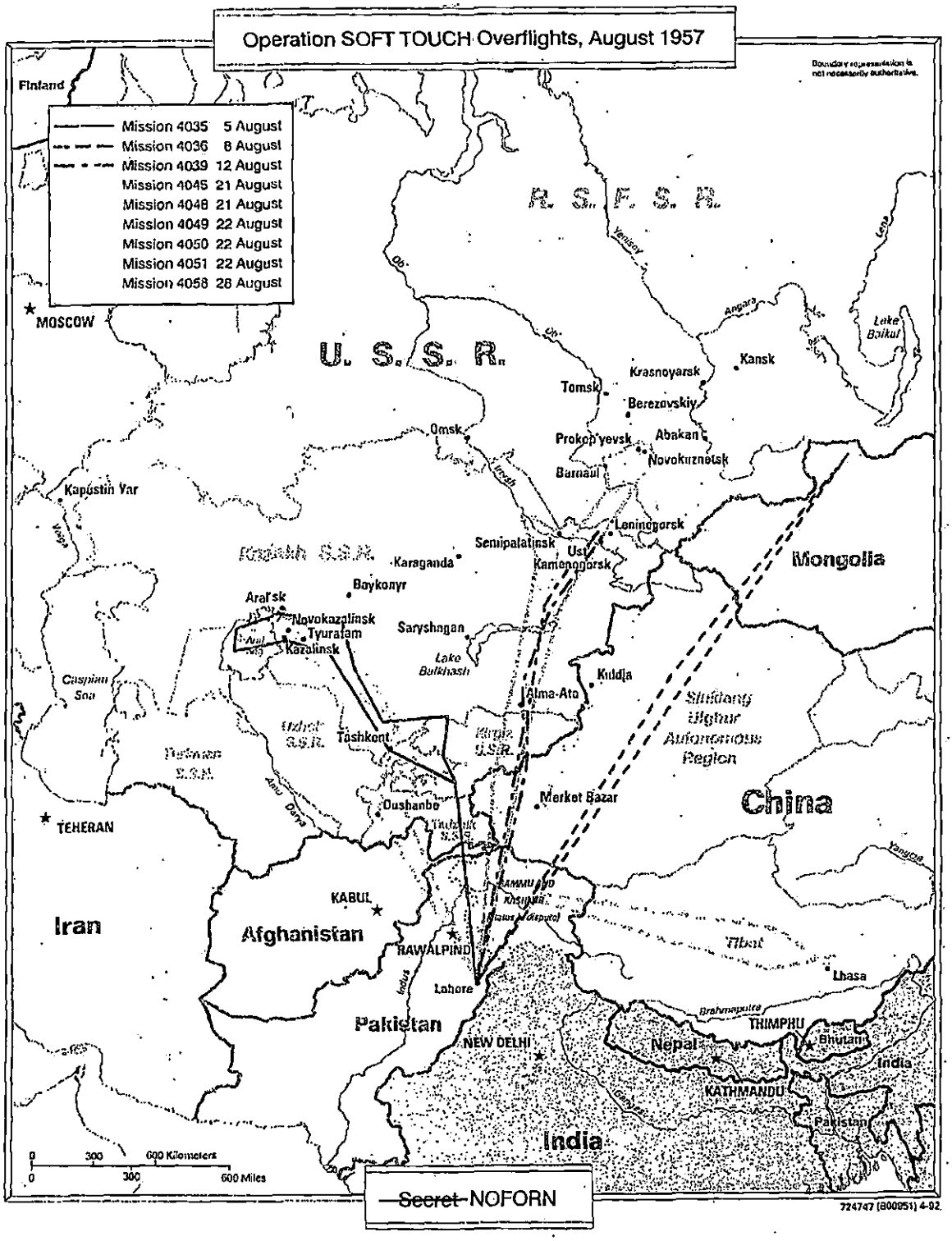
DETACHMENT B FLIGHTS FROM PAKISTAN

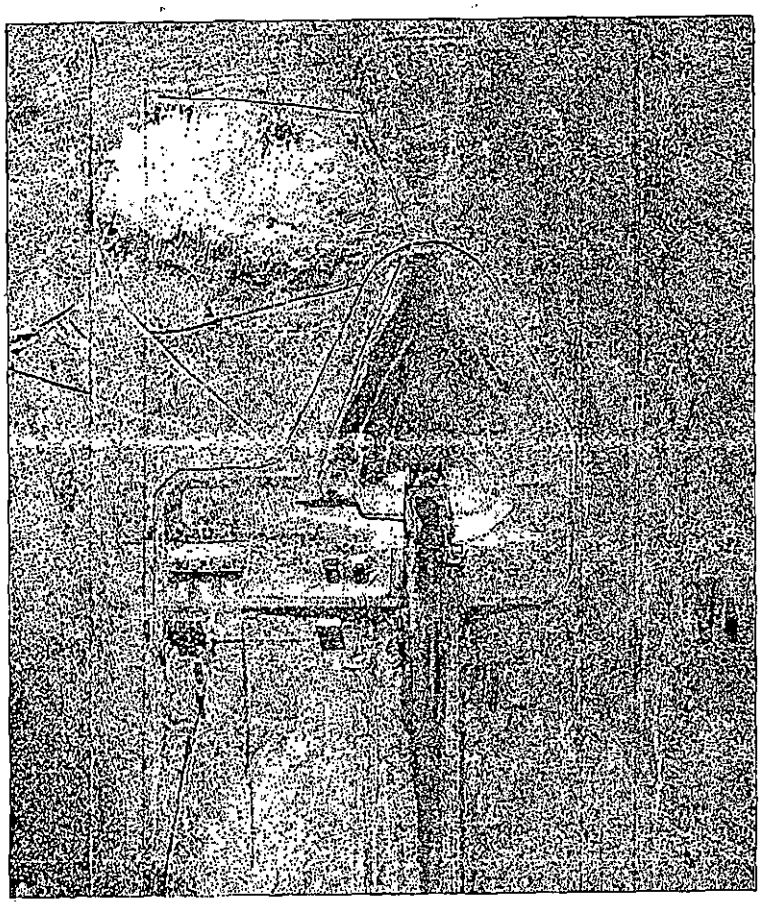
The most important series of overflights in the summer of 1957 were those that Detachment B staged to gather intelligence on the Soviet Union's guided missile and nuclear programs. President Eisenhower had approved these overflights at the meeting on 6 May 1957, provided that Pakistan allowed the U-2s to operate from its territory (the desired targets were too far away from the U-2 base in Turkey). [redacted] Richard Bissell's personal assistant, and [redacted] met with President Iskander Mirza, Prime Minister Huseyn Shaheed Suhrawardy, and Army Commander Gen. Ayub Khan between 3 and 7 June 1957 and received permission to operate from Lahore. The airfield at Peshawar, a more desirable location, was not available because of repair work. Detachment B at Ankara ferried four of its U-2s, two of which were dirty birds, to Lahore. A C-124 brought in eight pilots and ground crews to prepare for missions over the Soviet Union and the People's Republic of China (PRC) beginning on 4 August (Operation SOFT TOUCH). During a 23-day period, these aircraft made nine flights: seven over the USSR and two over the PRC. Although one of the seven flights over the USSR was a failure because the camera malfunctioned after taking only 125 exposures, the remaining missions over Central Asia were a complete success, producing a bonanza of information that kept scores of photointerpreters busy for more than a year.[82]

The 5 August flight, a dirty bird piloted by Buster Edens, was the first to photograph the major Soviet space launch facility east of the Aral Sea in Kazakhstan. None of the mission planners was certain just where the range was located, so the U-2 pilot followed the rail lines in the area. As a result, the plane did not pass directly over the rangehead and obtained only oblique photography.

Although known in the West today as Tyuratam, this missile installation had no name when it was first photographed in August 1957. In preparation for a briefing to President Eisenhower on the SOFT TOUCH photography, Dino Brugioni, an assistant to PID chief Arthur Lundahl, examined all the existing maps of the area to see if he could find a place name for the missile base. Only one map, made by the Germans during World War II, showed a community in the vicinity of the missile facility. The settlement's name was Tyuratam, which means "arrow burial ground" in the Kazakh language, and this

[82] *OSA History*, chap. 12, pp. 19-20 (~~TK~~ Codeword); *NPIC History*, vol. 1, pp. 159-161 (S).


Operation SOFT TOUCH Overflights, August 1957
Mission 4035 5 August
Mission 4036 8 August
Mission 4039 12 August
Mission 4045 21 August
Mission 4048 21 August
Mission 4049 22 August
Mission 4050 22 August
Mission 4051 22 August
Mission 4058 28 August
Finland
MOSCOW
U. S. S. R.
R. S. F. S. R.
Kansk
Krasnoyarsk
Tomsk
Berezovskiy
Omsk
Prokop'yevsk
Abakan
Novokuznetsk
Barnaul
Kapustin Yar
Leninogorsk
Semipalatinsk
Ust' Kamenogorsk
Mongolia
Karaganda
Baykonyr
Aral'sk
Novokazalinsk
Tyuratam
Kazalinsk
Saryshagan
Lake Balkhash
Kuldja
Alma-Ata
Caspian Sea
Tashkent
Dushanbe
Merket Bazar
China
TEHERAN
Iran
KABUL
Afghanistan
RAWALPINDI
Lahore
Pakistan
NEW DELHI
India
Nepal
KATHMANDU
THIMPHU
Bhutan
Lhasa
Tibet
Lake Baikal
0 300 600 Kilometers
0 300 600 Miles
~~Secret~~ NOFORN
724747 (800851) 4-92

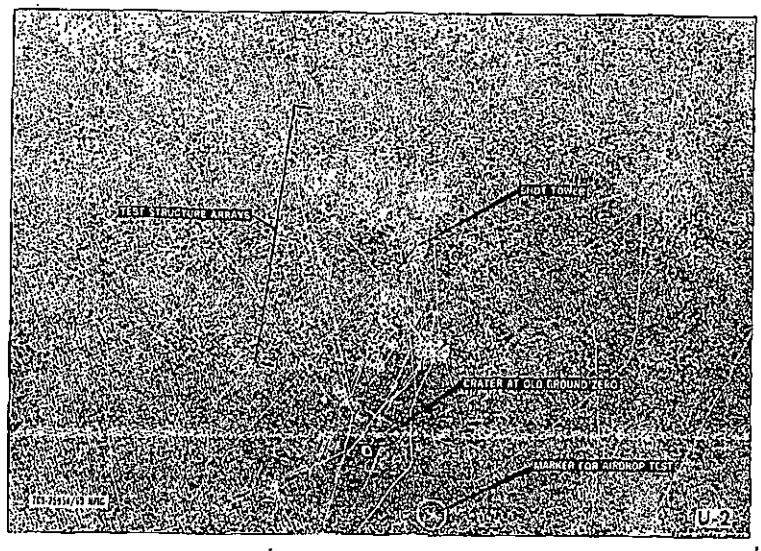


U-2 photography of Tyuratam Missile Testing Range

was the name Brugioni gave the missile base. Official Soviet releases concerning this base have always referred to it as Baykonur, but the community of Baykonyr is actually more than 200 miles north of Tyuratam.[83]

While PID was still analyzing the SOFT TOUCH photography, the Soviet Union announced the successful launch of an intercontinental ballistic missile (ICBM) from Baykonyr (Tyuratam). On 26 August 1957, the Soviet news agency TASS stated that a "super-long-range multistage intercontinental ballistic rocket" had been successfully tested, adding "it is now possible to send missiles to any

[83] Lundahl and Brugioni interview (TK Codeword).



Semipalatinsk Nuclear Weapons Proving Ground, 22 August 1957

part of the world."[84] The Soviet announcement made the intelligence community want even more information on Tyuratam, and a second U-2 piloted by Edwin K. Jones flew over the area on 28 August 1957, just one week after the Soviet ICBM launch. This mission obtained excellent vertical photographs of the main launch complex, and photointerpreters soon determined that the Soviets had only one launchpad at Tyuratam. The base was not photographed again until 9 July 1959, at which time it still had only one launch pad, although two more were under construction.[85]

On 20 and 21 August 1957, U-2s conducted the first overflights of the Soviet nuclear testing grounds at Semipalatinsk, north-northwest of Lake Balkhash. The first mission, piloted by Sammy V. C. Snider, passed over part of the proving grounds, flew on to Novokuznetsk, and then proceeded to Tomsk, where it began its return leg that included coverage of a very large uranium-processing facility at the new city of Berezovskiy. In the second mission, James Cherbonneaux flew directly over the Semipalatinsk proving grounds only four hours before a half-megaton device was detonated. In fact, the U-2 unknowingly photographed the aircraft that was to drop the

[84] "Is Russia Ahead in Missile Race," *US News and World Report*, 6 September 1957, pp. 30-33.

[85] Mission folders 4058 (28 August 1957) and 4125 (9 July 1959), OSA records, job 67-B-972, boxes 8 and 11 (~~TS~~ Codeword).

nuclear device. These photographs also revealed evidence of a recent, low-yield, above-ground nuclear test

25X1, E.O.13526

On its way to Semipalatinsk, the 21 August mission flew a search pattern over the western end of Lake Balkash looking for another Soviet missile-related installation and made the first photographs of what was later determined to be the new missile test center at Saryshagan. This facility was used to test radars against incoming missiles fired from Kapustin Yar, 1,400 miles to the west. Saryshagan later became the center for the development of the Soviet Union's advanced antiballistic missile (ABM) weapon system.

On 23 August 1957, DDCI Cabell, Richard Bissell, and Air Force Chief of Staff Twining met with President Eisenhower to report on the results of Operation SOFT TOUCH. They showed the President some of the photographic results of the earlier missions and reported on the effects of the antiradar measures. Although the antiradar measures had not proved successful, the photographic yield from the missions was extremely valuable. Bissell then informed the President that the SOFT TOUCH operation was just about to conclude with the transfer of the aircraft back to Adana. He asked permission for one of the U-2s to make another overflight of the Soviet Union on this return trip, but the President denied the request, not wishing to conduct any more overflights than were necessary.[87]

THE DECLINE OF DETACHMENT A

During the summer of 1957, all overflights of the Soviet Union were conducted by either Detachment B or Detachment C. Detachment A in Germany was a less desirable starting point for overflights of the Soviet Union because such missions had to cross Eastern Europe first, increasing the likelihood of detection and diplomatic protests. Furthermore, the Soviet Union's air defense and radar networks were strongest along its western borders, so Detachment B missions over the southern portion of the Soviet Union and Detachment C missions

[86] Mission folder 4045 (20 August 1957) and 4050 (21 August 1957), OSA records, job 67-B-972, box 8 (~~[illegible]~~ Codeword).

[87] Andrew J. Goodpaster, Memorandum for the Record, 23 August 1957, WHOSS, Alpha, DDEL. ~~(S)~~.

in the Far East were less risky than those conducted by Detachment A. Finally, the main target of U-2 photography after the bomber issue receded was Soviet missile and nuclear progress. The testing areas for these weapons were located in the vast open spaces of the south-central and eastern portions of the Soviet Union, which lay beyond the range of Detachment A's aircraft.

The decline in importance of Detachment A had begun with the President's standdown order of 10 July 1956. During the next three months, the detachment conducted only 11 missions, all over the Mediterranean region rather than the original target of the Soviet Union, and the slow pace of activity and change in mission adversely affected pilot morale. One of the detachment's aircraft was lost in a crash on 17 September, killing pilot Howard Carey and garnering unwanted publicity. Conditions improved when the detachment moved to the newly renovated facility at Giebelstadt in early October 1956, but security now became a problem there. Detachment A personnel discovered that a long, black Soviet-Bloc limousine was parked at the end of the Giebelstadt runway whenever the U-2s took off.[88]

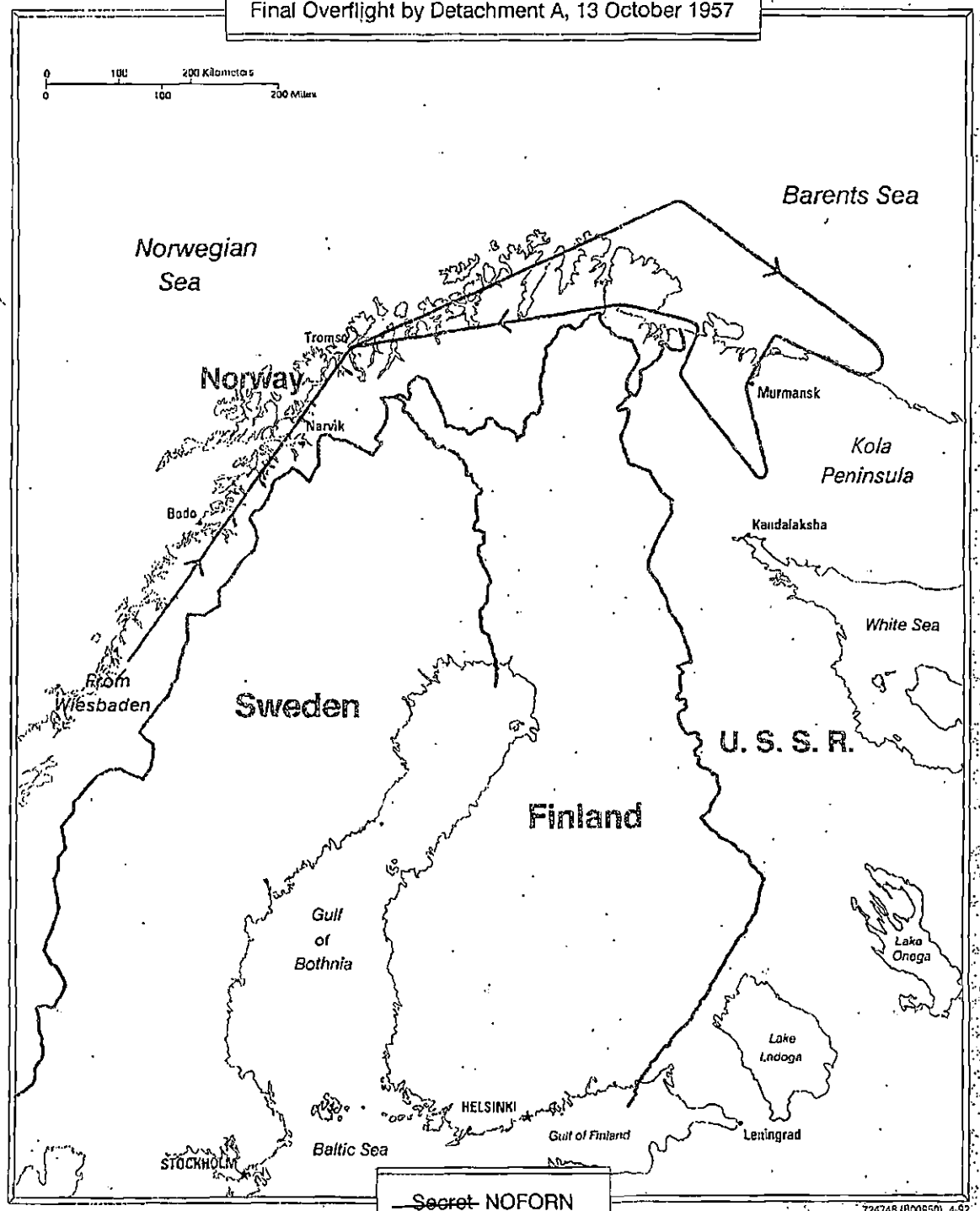
During the next year, Detachment A mounted only four overflights. The first two were over Eastern Europe: one over Bulgaria on 10 December 1956 and the other over Albania on 25 April 1957. Then a long period of inactivity followed, ending with a third mission on 11 October 1957, which conducted electronic surveillance of Soviet naval maneuvers in the Barents Sea. The final overflight of Detachment A, mission 2040 on 13 October 1957, flew north over Norway to the Bering Sea, turned southeast to overfly Murmansk, and then exited to the north, returning to Germany via Norway.[89]

Although the final missions of Detachment A achieved excellent results, project headquarters had already decided that Western Europe was not a satisfactory location for overflights of the Soviet Union and had notified Detachment A on 20 September 1957 that its operations would cease in November. By 15 November 1957, all of the detachment's personnel and aircraft had returned to the United States. During Detachment A's 17-month period of operations, seven pilots

[88] *OSA History*, chap. 11, pp. 41-42 (~~CPS~~ Codeword).

[89] Mission folders 4018 (10 December 1956), 2036 (25 April 1957), 2037 (11 October 1957), 2040 (13 October 1957), OSA records; job 67-B-972, box 7, and job 67-B-328, box 6 (~~CPS~~ Codeword).

Final Overflight by Detachment A, 13 October 1957



~~Secret~~ NOFORN

724748 (B00850), 4-92

had flown a total of 23 missions: six over the Soviet Union, five over Eastern Europe, and most of the remaining 12 missions over the Mediterranean area.[90]

COOPERATION WITH NORWAY

The final missions of Detachment A had one unforeseen result: the beginning of cooperation between CIA and the Norwegian Intelligence Service on the U-2 program. Norwegian radars tracked the overflights of the Barents Sea and Murmansk in October 1957, and Col. Vilhelm Evang, Chief of the Norwegian Defense Staff Intelligence Service, 25X1, E.O.13526 In March 1958, Colonel Evang came to Washington and received a briefing from Arthur Lundahl together with sanitized photos that did not reveal the altitude of the aircraft or the focal length of the camera. The Agency provided additional photos to Norway during a visit by a PID staff member in July.[91]

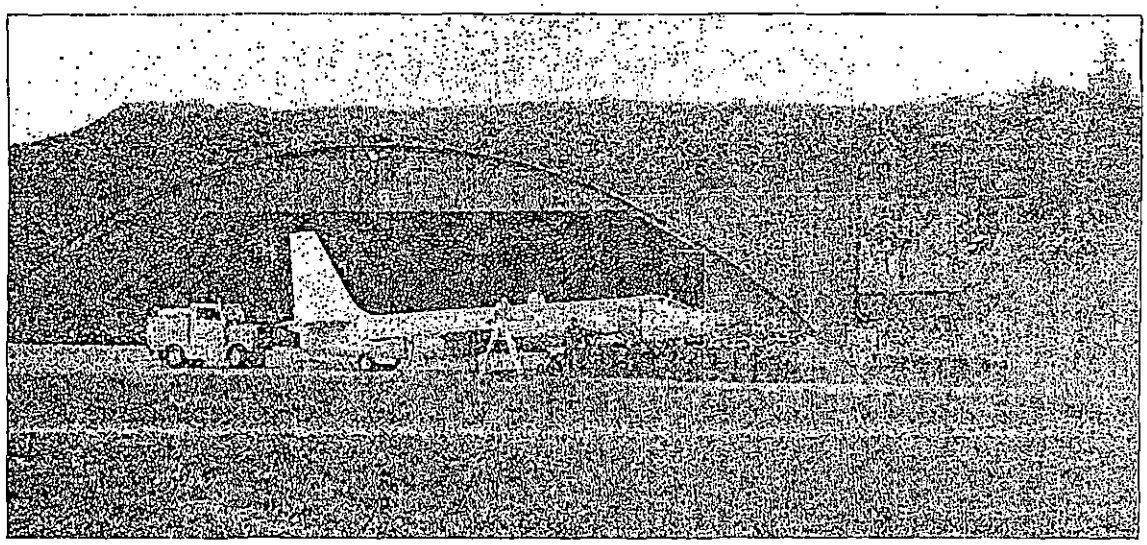
Later that year Norway agreed to provide an airfield for the United States to conduct U-2 flights that did not violate Soviet airspace. On 15 September 1958, a Detachment B staging party arrived at Bodo Air Force Base, 25X1, E.O.13526 this was followed by an ELINT collection flight over international waters (the Kara Sea) on 25 October. On 6 November, the U-2 returned to Adana by conducting a lengthy ELINT collection flight along the borders of the Soviet Union and East Germany. During the initial portion of the mission, when the U-2 flew along the Soviet-Finnish border and then turned east over the Gulf of Finland to come within 60 miles of Leningrad (while remaining over international waters), the aircraft was the target of 23 unsuccessful Soviet intercept attempts.[92]

---

[90] *OSA History*, chap. 11, p. 44; chap. 19, annex 120 (TS Codeword).

[91] Ibid., chap. 11, pp. 44-45 (TS Codeword); *NPIC History*, vol. 3, pp. 447-8 (S).

[92] Mission folders 1482 (9 October 1958), 4092 (25 October 1958), and 4093 (6 November 1958), OSA records, job 67-B-972, boxes 3 and 10.



U-2 at Bodo, Norway

DECLINING OVERFLIGHT ACTIVITY

Operation SOFT TOUCH (4-27 August 1957) proved to be the high water mark of U-2 operations against the Soviet Union. Detachment B staged one more overflight on 10 September 1957, when a U-2 piloted by William Hall flew from Adana to photograph the Kapustin Yar Missile Test Range for the first time since the RAF's overflight in 1953, obtaining photographs of a large medium-range ballistic missile (MRBM) on the launchpad. Six days later Detachment C conducted its successful overflight of the ICBM impact site at Klyuchi, and October saw the final two overflights of Detachment A. After these missions, penetration overflights became a rarity. There would be only six more during the next 32 months: one, in 1958; two, in 1959; and three, in 1960 (one of which was unsuccessful). During this period, President Eisenhower did authorize a number of flights along Soviet border areas that occasionally penetrated short distances inside the border, but the Chief Executive had become extremely wary of authorizing "deep penetration" overflights, which invariably brought protests from Moscow.

The border flights took place under tight controls. Beginning in the fall of 1957, all messages from Washington to Adana giving coordinates for flights along the Soviet border contained the statement: "This is not a penetration overflight" and warned about flying too close to Soviet borders. The Soviets even attempted to shoot down

U-2s flying well within international airspace above the Black Sea, as was the case on 27 October 1957, when electronic intelligence equipment on a U-2 flight over the Black Sea that never violated Soviet airspace revealed 12 attempts at interception by Soviet fighters.[93]

The sole U-2 overflight of 1958 was conducted by a dirty bird from Detachment C. On 1 March 1958, mission 6011 overflew the Soviet Far East and photographed the Trans-Siberian Railroad, Sovetskaya Gavan', the Tatar Strait, and a strange installation at Malaya Sazanka, which was eventually determined to be a structure for mating nuclear devices with their detonators. This was the first and only U-2 overflight of the Soviet Union staged from Japan.[94]

On 5 March 1958, the Soviet Union delivered a vigorous protest concerning this mission, prompting President Eisenhower to tell Colonel Goodpaster on 7 March to inform the CIA that U-2 flights were to be "discontinued, effective at once."[95] This standdown was to last more than 16 months, until July 1959. The Soviets had not been fooled by the antiradar devices carried by mission 6011, as was demonstrated by the detailed information about the mission contained in a Soviet aide-memoire delivered on 21 April 1958. It was clear that dirty bird aircraft were not effective and that Soviet radar operators had little difficulty in tracking them. At this point, the Agency abandoned the use of the antiradar devices on the U-2. As a substitute, Lockheed began working to develop a paint with radar-suppressant qualities, but this project also proved unsuccessful.

The U-2s were not the only cause for the Soviet protests that so vexed the President. On 27 June 1958, an Agency C-118 on a courier flight from Adana to Teheran strayed into Soviet Armenia and was shot down; the Soviets captured two survivors, including one Agency employee. Ten days later the Air Force began launching balloons designed to fly across the Soviet Union and Eastern Europe. This new balloon project (known as WS-461L) had been authorized by President Eisenhower on 25 June after Deputy Secretary of Defense Donald Quarles argued that a small number of balloons should be

[93] Mission folder 4061 (27 October 1957), OSA records, job 67-B-972, box 8 (~~CPS~~ Codeword).

[94] Mission folder 6011 (1 March 1958), OSA records, job 67-B-972, box 13 (~~CPS~~ Codeword).

[95] Andrew J. Goodpaster, Memorandum for the Record, 7 March 1958, WHOSS, Alpha, DDEL (~~CPS~~, declassified).

launched to take advantage of a newly discovered change in the west-to-east jet stream. Normally, this fast-moving air current stayed at an altitude of 55,000 feet, but, during June and July, it turned abruptly upward over the Bering Sea just west of Alaska, climbed to 110,000 feet, and then reversed direction. One of the key arguments that convinced the President to approve the project was Quarles's claim that the balloons' "chance of being detected is rather small and their identification or shootdown practically nil."[96]

Release of the balloons took place from an aircraft carrier in the Bering Sea on 7 July 1958. Nothing was heard about them until 28 July, when Poland sent a note protesting the overflight of a US-made, camera-carrying balloon that had fallen to earth in central Poland. The loss of this balloon was because of human error. Each balloon was equipped with a timing device that would cause it to drop its camera and film payload after crossing the target areas. An Air Force technician aboard the aircraft carrier had calculated that the balloons should cross the Eurasian landmass in about 16 days. Thus, he adjusted regulators aboard the balloons to cause automatic descent after 400 hours aloft. When bad weather delayed the launch for three successive days, however, the technician forgot to reset the timing devices. As a result, one payload fell into Poland. None of the three WS-461L balloon payloads was recovered.[97]

The Polish protest was quickly followed by a Soviet note protesting the balloons' violation of the Soviet Union's airspace. Several months later, the Soviets placed the US balloon and photographic equipment on display in Moscow for the world's press. President Eisenhower was angry that the Defense Department's assurances that the balloons would not be detected had proved false. Even worse, one of the balloons had been recovered by the Poles because the Air Force had disobeyed his instructions for the balloon project. When the Air Force had proposed the use of timers to bring down the balloons at the end of the mission, Eisenhower had said no, fearing that a malfunction could cause the balloons to come down prematurely. Furious at the Air Force's insubordination, the President ordered General

[96] Andrew J. Goodpaster, Memorandum for the Record, 25 June 1958, WHOSS, Alpha, DDEL. ~~(U)~~.

[97] Donald E. Welzenbach, "Observation Balloons and Weather Satellites," *Studies in Intelligence* 30 (Spring 1986): pp. 26-28 ~~(S)~~.

Equipment from a WS-461L balloon on display in Moscow, 11 October 1958

Goodpaster on 29 July 1958 to tell the Air Force that "the project is to be discontinued at once and every cent that has been made available as part of any project involving crossing the Iron Curtain is to be impounded and no further expenditures are to be made."[98]

Two days later Eisenhower followed up this order with a formal memorandum to Secretary of Defense Neil McElroy telling him that "there is disturbing evidence of a deterioration in the processes of discipline and responsibility within the armed forces." He cited, in particular, "unauthorized decisions which have apparently resulted in certain balloons falling within the territory of the Communist Bloc" and overflights over routes "that contravened my standing orders."[99]

On 2 September 1958, there was another violation of Soviet airspace when an unarmed Air Force EC-130 on an electronic intelligence collection mission crossed from Turkey into Soviet Armenia and was shot down by Soviet fighter aircraft. Six of the men on board were killed and the remaining 11 were never heard from again, despite State Department attempts to get the Soviet Union to reveal their fate.[100]

---

[98] Andrew J. Goodpaster, Memorandum for the Record, 29 July 1958, WHOSS, Alpha, DDEL (S); Goodpaster interview ~~(S)~~.

[99] Quoted in Ambrose, *Eisenhower: The President*, pp. 475-476.

[100] "US Representations to the Soviet Government on C-130 Transport Shot Down by Soviet Fighter Aircraft," *US Department of State Bulletin*, 23 February 1959, pp. 262-271; Beschloss, *Mayday*, p. 159.

President Eisenhower was disturbed by the increased superpower tension that had resulted from violations of Soviet airspace by US balloons and aircraft because he still hoped to enter into arms limitation negotiations with the Soviets. On 8 September 1958, the United States sent a note to the Soviet Union calling for a Soviet answer to US proposals for a "study of the technical aspects of safeguards against the possibility of surprise attack." One week later the Soviets agreed to participate and suggested that the talks begin in Geneva on 10 November 1958. President Eisenhower was also attempting to persuade the Soviet Union to begin talks aimed at eliminating the atmospheric testing of nuclear weapons. These efforts began with a 22 August 1958 offer to suspend US nuclear tests for one year on the condition that the Soviet Union also refrain from further tests and join in negotiations. On 30 August, Soviet Premier Nikita Khrushchev accepted the proposal and agreed to start talks on 31 October 1958 in Geneva. When the talks began, however, the Soviets refused to agree to a test ban and carried out nuclear tests at Semipalatinsk on 1 and 3 November. Nevertheless, during the late summer and early autumn of 1958, President Eisenhower, determined to reduce to a minimum any aggravation of the Soviets, kept the U-2 overflight program in limbo.[101]

In November 1958, relations with the Soviet Union worsened after Khrushchev precipitated a new crisis over West Berlin by announcing plans to sign a peace treaty with East Germany by May 1959. He stated that such a treaty would terminate Allied rights in West Berlin. Four days later, Soviet troops began harassing US Army truck convoys on the highways leading from West Germany to West Berlin. Although this new Berlin crisis never became as threatening as the blockade of 1948-49, President Eisenhower wished to avoid any actions that would provoke the Soviets. Tension over West Berlin was, therefore, an additional reason for continuing to keep the U-2 away from the Soviet Bloc.[102]

CONCERNS ABOUT SOVIET COUNTERMEASURES AGAINST THE U-2

Another reason for President Eisenhower's growing reluctance to authorize flights over the Soviet Union may have been concern that the Soviets were developing countermeasures that would enable them to

[101] Ambrose, *Eisenhower: The President*, pp. 489-491.

[102] Ibid., pp. 502-504.

shoot down a U-2. Before the program started, Richard Bissell had estimated that the U-2 would be able to fly over the Soviet Union with impunity for only about two years. This period was already over, and the Soviets were working frantically to devise a means to stop U-2 overflights. From the very beginning, Soviet air defense units had not only tracked U-2s with radars, but had also made repeated efforts to shoot them down with antiaircraft weapons and interceptor aircraft. In 1956 such attempted interceptions had involved primarily MiG-15s and MiG-17s, which could barely reach 55,000 feet. The advent of MiG-19s and MiG-21s, which could climb even higher, provided a greater threat for U-2 pilots.

Realistic training for pilots learning to intercept the U-2 became possible after the Soviets developed a new high-altitude aircraft, the Mandrake, which was actually an improved version of the Yakovlev-25 all-weather interceptor. The Mandrake used a high-lift, low-drag wing design similar to that employed by the U-2, but its twin engines made it heavier. The Mandrake's operating altitude was 55,000 to 65,000 feet, and its maximum altitude was 69,000, far less than the 75,000 feet reached by the U-2. Like the U-2, the Mandrake's wings would not tolerate great stresses, so it could not be used as an attack aircraft at the high altitudes at which both planes operated. Between 1957 and 1959, Yakovlev built 15 to 20 of these aircraft in two versions: the Mandrake-R or YAK-25RM and the Mandrake-T, sometimes called the YAK-26. These high-altitude aircraft were used to overfly the Middle East, India, China, and Pakistan, as well as border regions of NATO nations in Europe during the late 1950s and early 1960s. It is not believed that Mandrakes ever attempted to overfly the continental United States.[103]

Beginning in late 1957, the Mandrake served as a practice target for pilots of high-performance Soviet MiG-19 and MiG-21 interceptors. The Soviet technique that most concerned U-2 pilots was the "snap up" or power dive and zoom climb. In this maneuver, ground-based radar operators would direct the interceptor aircraft along the same flight path as the U-2. When the MiG pilot achieved the same compass heading as the U-2 flying more than 10,000 feet above him, he would put his aircraft into a shallow dive to pick up

[103] "Yakovlev Yak-25RM Mandrake," *Jane's Defence Weekly*, vol. 3, no. 7, 16 February 1985.

speed, apply full throttle to the engine, then pull back on the stick and zoom as high as he could. In this manner the Soviet pilot hoped to come up directly beneath the U-2 so he could use his guns and missiles against the shiny U-2 etched in silver against the dark blue-black of space. Using this maneuver, some MiGs were able to climb as high as the U-2 but seldom got very close. At this height the MiGs were completely out of control; their small, swept-back wings provided insufficient lift; and their control surfaces were too small to maintain aircraft stability. U-2 pilots often spotted MiGs that reached the apex of their zoom climbs and then fell away toward the earth. The US pilots' greatest fear was that one of the MiGs would actually collide with a U-2 during a zoom climb.[104]

U-2 pilots complained that they felt like ducks in a shooting gallery under these circumstances and suggested that the underside of the silvery aircraft be camouflaged in some manner. Kelly Johnson had originally believed the U-2 would fly so high that it would be invisible, thus eliminating the need to paint the aircraft and thereby avoiding the added weight and drag that paint produced. The paint penalty was calculated to be a foot of altitude for every pound of paint. A full coat of paint cost the U-2 250 feet of altitude, substantially less than the 1,500-foot penalty paid for the addition of dirty bird devices.

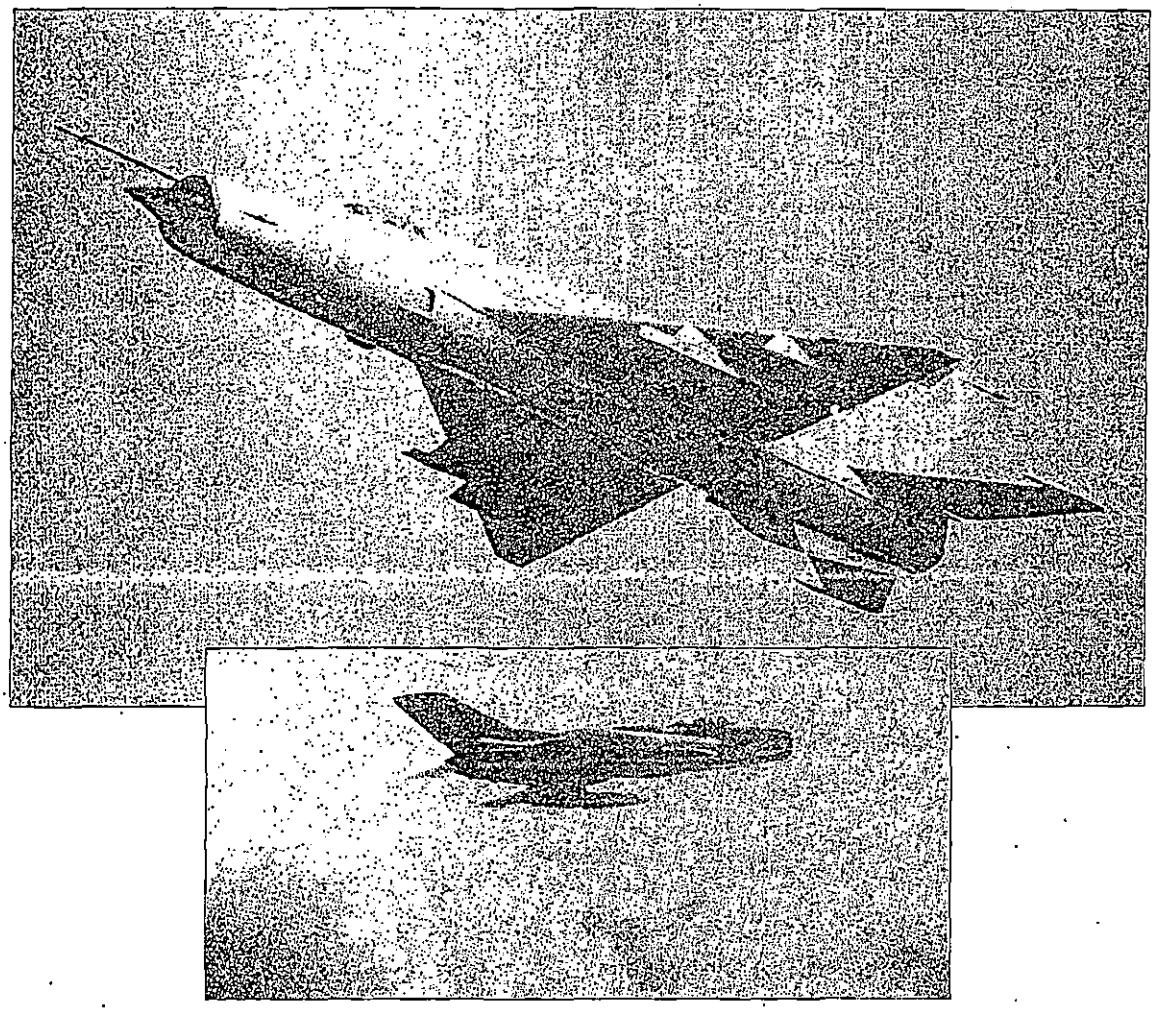
By late 1957, Johnson agreed that something had to be done. After a series of tests over Edwards AFB, Lockheed began coating the U-2s with a standard blue-black military specification paint on top and a lighter cloud-blue paint below. Subsequent tests over Nevada revealed that the U-2s were less conspicuous when painted all over with a matte-finish blue-black color, which helped them blend with the dark canopy of space.[105]

MORE POWERFUL ENGINES FOR THE U-2

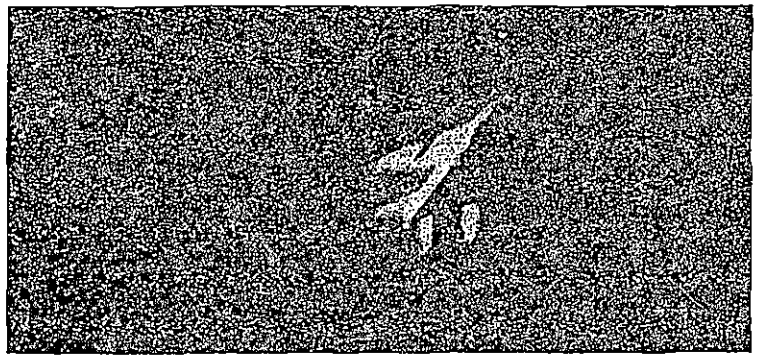
Less conspicuous paints were not the only answer to the growing threat of Soviet interceptors. A more powerful engine would increase the U-2's maximum altitude, which was the surest way to protect the aircraft from all Soviet threats. During late 1958 and early 1959, Lockheed began refitting the Agency's 13 remaining U-2s—originally the Agency had taken delivery of 20 planes and the Air

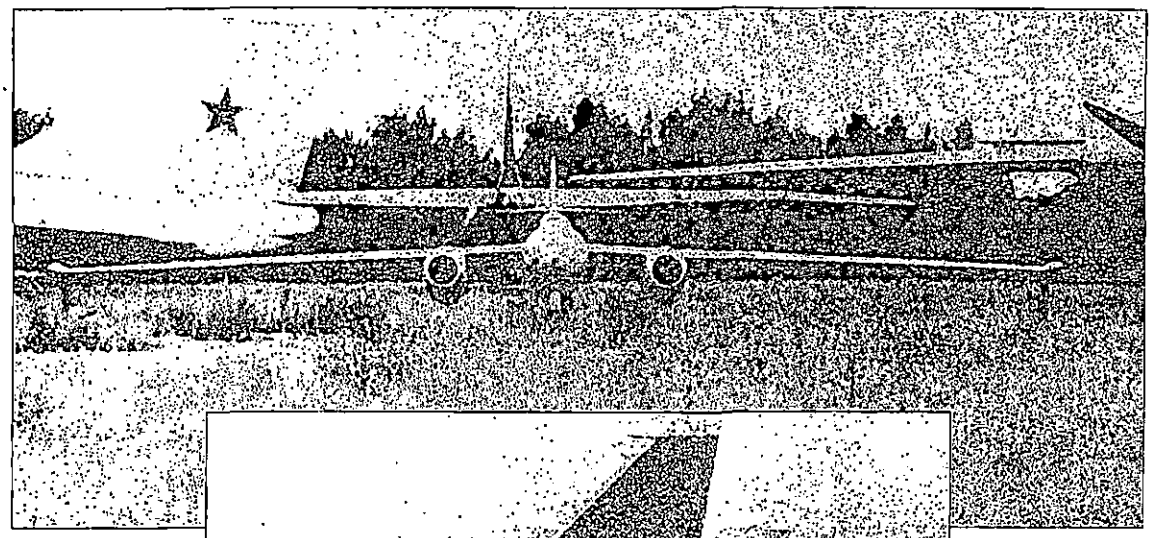
[104] Information supplied by Jacob Kratt and James Cherbonneaux to Donald E. Welzenbach, May 1986.

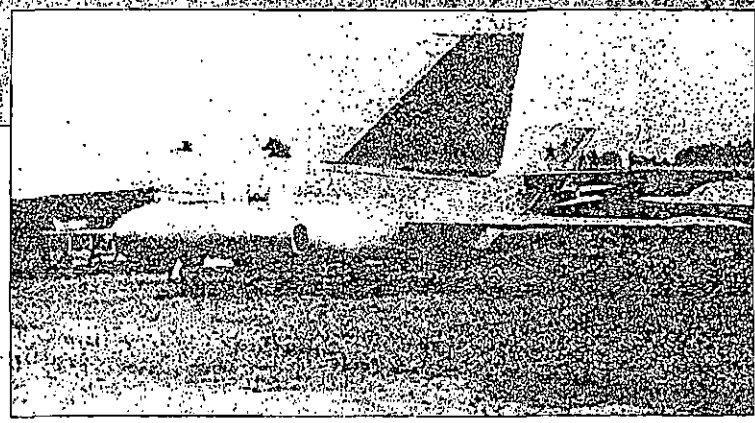
[105] Lockheed contracts, OSA records (S).

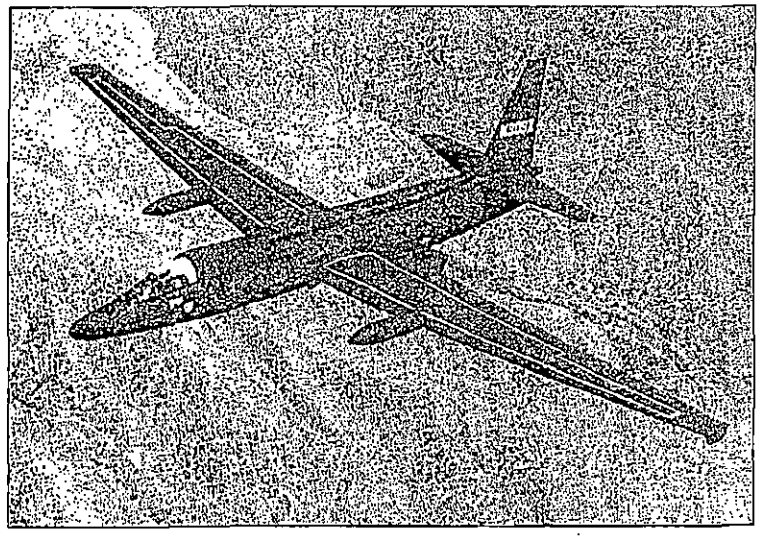


Soviet MiG-21 interceptor (top), Soviet MiG-19 interceptor (middle), Soviet MiG-19 photographed by a U-2, 13 October 1957 (bottom)









YAK-25RD Mandrake on display at the Gagarin Military Academy Museum (top and middle)
U-2 in the new black paint scheme (left)

Force of 31—with the more powerful Pratt & Whitney J75-P13 jet engine. This new power plant generated 4,200 pounds more thrust while adding only 2,050 pounds more weight. With its greater power, the engine permitted the U-2 to reach operational altitude more quickly, thereby reducing the telltale contrails that the U-2 produced as it passed through the tropopause at 45,000 to 55,000 feet. With the new engine, U-2 passed through this portion of the atmosphere faster and did so before entering hostile airspace, thus reducing the chance of visual detection. The J75 power plant also made it possible for the U-2 to carry a larger payload and gain another 2,500 feet in altitude, permitting it to cruise at 74,600 feet. The new engines were in very short supply because of the needs of the Air Force's F-105 construction program, but Colonel Geary used his Air Force contacts to obtain an initial supply of 12 engines. The Air Force never equipped its original U-2s with the J75 engines.[106]

Detachment C in Japan received the first of these re-engined aircraft, known as U-2Cs, in July 1959, and two more arrived in Turkey for Detachment B in August. All Agency U-2s had the new engines by the summer of 1962, but by then only seven CIA U-2s remained in service.

INTERVENTION IN LEBANON, 1958

Although the U-2 was used less and less for its original role of gathering strategic intelligence on the Soviet Bloc, it had acquired the new mission of providing US decisionmakers with up-to-date information on crisis situations all around the world. The first use of the U-2 to gather tactical intelligence occurred during the 1956 Suez Crisis. Afterward, U-2s from the Turkish-based Detachment B conducted periodic overflights to monitor the situation in the troubled Middle East, and they became especially active during the summer of 1958.

On 15 July 1958, President Eisenhower ordered US troops to land in Lebanon in response to a request for assistance by Lebanese President Camille Chamoun. Three months earlier, Eisenhower had turned down a similar request because the rioting that had led President Chamoun to ask for American aid had died down before intervention became necessary. In July, however, President Eisenhower saw the overall situation in the Middle East as much more threatening. On 14 July forces aligned toward Egyptian President Gamal

[106] *OSA History*, chap. 16, p. 8 (TS Codeword); Geary interview (S).

Abdel Nasser overthrew the Government of Iraq and assassinated the royal family. Long concerned by the growing influence of Nasser, who had close ties to the Soviet Union and now headed both Egypt and Syria in the new United Arab Republic, President Eisenhower decided that US intervention was necessary to stabilize the situation in Lebanon and to show Nasser that the United States was willing to use force to defend its vital interests in the region. Before intervening in Lebanon, the United States consulted with the United Kingdom, which also decided to intervene in the Middle East by sending paratroopers to assist the Government of Jordan on 17 July.

With US Marines and Army troops deployed in a potentially hostile situation in Lebanon, US military commanders and intelligence community analysts immediately requested tactical reconnaissance flights to look for threats to the US units and evidence that other Middle Eastern countries or the Soviet Union might be preparing to intervene. The U-2s of Detachment B in Turkey carried out these missions.

Because tactical reconnaissance required an immediate readout of the films taken, the Photographic Intelligence Center (the new name for the Photo-Intelligence Division from August 1958) quickly reopened the film-developing unit at Adana and staffed it with lab technicians and photointerpreters. Throughout the summer of 1958, Detachment B U-2s brought back photography of military camps, airfields, and ports of those Mediterranean countries receiving Soviet arms. The detachment also kept a close watch on Egyptian-based Soviet submarines, which posed a threat to US 6th Fleet ships in the Mediterranean. In addition, U-2s flew occasional electronic intelligence collection missions along the Soviet border and over the Black Sea without entering Soviet airspace. In late August, as the crisis in the Middle East eased, the United States began withdrawing its 14,300 troops. It was not until 25 October, however, that the last American soldier left Lebanon.[107]

BRITISH PARTICIPATION IN THE U-2 PROJECT

Shortly after the withdrawal of US troops from Lebanon, a new group of pilots joined the U-2 project. In November 1958, four RAF officers arrived at Detachment B at Adana, thus beginning the United Kingdom's participation in U-2 operations. The British had first

[107] Ambrose, *Eisenhower: The President*, pp. 462-473.

become involved in the U-2 project in September 1956, when the United States supplied them with photography from U-2 missions. To handle U-2 material, the British created a new control system, which later merged with the US control system. By 1957 cooperation between the United Kingdom and the United States had expanded to include frequent consultation between the requirements and photo-interpretation organizations of both countries. James Reber and Arthur Lundahl made periodic trips to the United Kingdom for discussions with Alan Crick's UK Requirements Committee (generally known as the Crick Committee, later as the Joint Priorities Committee), the Joint Intelligence Committee, the Joint Air Reconnaissance Intelligence Center, and MI-6.[108]

The idea of using British pilots in the U-2 program first arose in the spring of 1957, when Richard Bissell—upset that his aircraft had not been allowed to fly over the Soviet Union since the December 1956 standdown—was searching for ways to reduce the political risks of overflights and thus obtain more frequent authorization for missions over the Soviet Union. One of his proposals was to use non-US pilots—possibly British—to increase the possibility of plausible denial in the event of a loss. At a meeting with key CIA, Defense Department, and State Department officials on 6 May 1957, President Eisenhower approved the concept of British participation in the U-2 project.[109]

During the next six months, Dulles and Bissell met with Sir Dick White, head of MI-6, and Air Vice Marshal William M. L. MacDonald, Assistant Chief of the Air Staff for Intelligence, on several occasions to discuss the proposal in general terms. At first the CIA did not push the proposal too hard because at the same meeting in which he approved British participation, President Eisenhower had consented to the resumption of U-2 missions over the Soviet Union, resulting in 10 overflights during the summer and early fall of 1957. But when flights ceased in October, the thought of British participation became more attractive. By early 1958, Bissell was pressing the British to begin training pilots in the U-2 even though no final political decision on their participation had been made. On 7 February 1958, Bissell instructed the [redacted] to ask Air Vice Marshal MacDonald if the RAF was prepared to

25X1, E.O.13526

[108] See, for example, the trip reports of the Ad Hoc Requirements Committee in the COMIREX records, IC Staff (TS Codeword).

[109] Andrew J. Goodpaster, Memorandum of Conference with the President, 6 May 1957 (TS); idem, "Record of Action—Meeting of May 6, 1957," WHOSS, Alpha, DDEL (TS).

select a group of pilots for the U-2 project. MacDonald agreed to Bissell's proposal and began recruiting RAF pilots to fly the U-2.[110]

In June 1958, representatives from the British Air Ministry came to project headquarters for an orientation and then sat down with CIA officials to work out an agreement on plans and procedures for the joint project. The two sides decided to establish a small RAF contingent that would be integrated into and supported by Detachment B at Adana. The British missions would be operationally controlled by CIA project headquarters. Soon afterward four British pilots began training in Texas. One of these pilots, Squadron Leader Christopher H. Walker, died in a training accident in July 1958. Because of the addition of RAF officers to the program, Project AQUATONE received a new codename, CHALICE. By the end of November 1958, three RAF pilots and a flight surgeon joined Detachment B at Adana with Turkish approval.

Formal approval by the political leaders of the United Kingdom and the United States had come several months earlier. On 27 August 1958, Prime Minister Harold Macmillan gave his approval to British participation in the project as long as he had the right to approve or disapprove all operational flights by RAF pilots. On the same day, President Eisenhower gave his approval in principle for the joint project.[111]

25X1, E.O.13526

Both sides stood to gain from the joint nature of the U-2 project. For Richard Bissell, British participation was a means to gain an additional source of authorization for overflights of the Soviet Union. Six months earlier, Bissell had cabled to [redacted] that he wanted British participation because it would "facilitate operations by them at times or under circumstances beyond the scope of authority accorded by US political authorities."[112]

President Eisenhower viewed British participation as a way to confuse the Soviets as to the sponsorship of particular overflights and also to spread the risk in the event of a loss. Furthermore, he was used

[110] *OSA History*, chap. 13, pp. 1-2 (~~TS~~ Codeword).

[111] Ibid., pp. 5-6 (~~TS~~ Codeword).

[112] Ibid., p. 1 (~~TS~~ Codeword).

to working closely with the British from his wartime experience and believed that their involvement in the U-2 program was a natural aspect of their alliance with the United States.[113]

On the British side, participation in U-2 flights was a logical extension of the close cooperation that already existed between the two countries on the U-2 program. The direct involvement of the British also enabled them to conduct additional flights in areas such as the Middle East that were of more intelligence interest to the United Kingdom than to the United States. The British also may have reasoned that direct participation in the program was the best way to ensure that they had a right to share in the U-2's take. Otherwise, the United States might decide at some point to cut off the flow of U-2 photography, as it had done during the 1956 Suez Crisis.

By November 1958, British pilots had joined Detachment B, and arrangements had been made for the title to the aircraft they would be using to be transferred on paper to the British Government. In a final exchange of letters between President Eisenhower and Prime Minister Macmillan in December, the President summarized the lines of authority for the joint program: "British missions are carried out on your authority and are your responsibility just as our activities are authorized and controlled here in accordance with the procedures I have established. In this sense, it could be said that we are carrying out two complementary programs rather than a joint one."[114]

Richard Bissell had achieved his goal of gaining another source of approval for overflights of the Soviet Union. In late 1959 and early 1960, this arrangement proved its value when British pilots conducted two highly successful missions over Soviet missile testing facilities at a time when President Eisenhower had not authorized an overflight for almost six months (see chapter 4). Most flights by the RAF pilots in Detachment B, however, took place in the Middle East, where the United Kingdom carried out 27 missions during the two years its pilots took part in overflights.

In Britain, the cover story for the RAF participation in the U-2 program was that British fliers were being trained to fly high-altitude weather-sampling missions for the RAF weather service. To support this cover, a U-2 was ferried to Watton RAF Base, England, in early

[113] Goodpaster interview ~~(S)~~.

[114] *OSA History*, chap. 13, pp. 10-11 (TS Codeword).

May 1959, and used to fly weather missions on 7 and 8 May before returning to Adana. Two more weather-sampling flights took place over England on 5 and 6 October 1959.

THE U-2 PROJECT AT THE BEGINNING OF 1959

Early 1959 saw Detachment B aircraft active primarily over Middle Eastern countries, with occasional overflights of Albania to check for reported Soviet missile installations. Detachment C mainly collected high-altitude weather data, although it also flew two missions over Tibet and Southwest China (see chapter 5). The overflight program against the Soviet Union seemed to be at a standstill, but pressures within the government were building to resume deep-penetration flights to resolve the growing "missile-gap" controversy.

Organizationally, the U-2 project underwent a major change after Richard Bissell became CIA's Deputy Director for Plans on 1 January 1959. At first glance, Bissell's selection seems unusual because he had spent most of his Agency career heading the U-2 project, but his first major assignment had been coordinating support for the operation that overthrew the leftist Government of Guatemala in 1954. Furthermore, Bissell's U-2 project was the major covert collector of intelligence against the CIA's primary target, the Soviet Union.

During his years as head of the Development Projects Staff (DPS), Bissell had opposed proposals to bring all Agency air activities together into a single office, fearing that he would lose control of the U-2 project. Once he became Deputy Director for Plans, his viewpoint changed; he was now in a position to consolidate all air activities under his own control. On 16 February 1959, the DPS became the Development Projects Division (DPD) of the Directorate of Plans (at the time known as the Deputy Directorate/Plans or DDP). Despite the tremendous increase in the scope of his duties after assuming control of the DDP, Bissell retained personal control of his previous Development Projects Staff projects: the U-2 program, another project to develop a photosatellite, and a third project to design a follow-on aircraft for the U-2 (OXCART). Although the amalgamation of all Agency air operations and the transfer of the U-2 project to the DDP made sense, the question remained as to whether one individual could effectively control all these different activities.

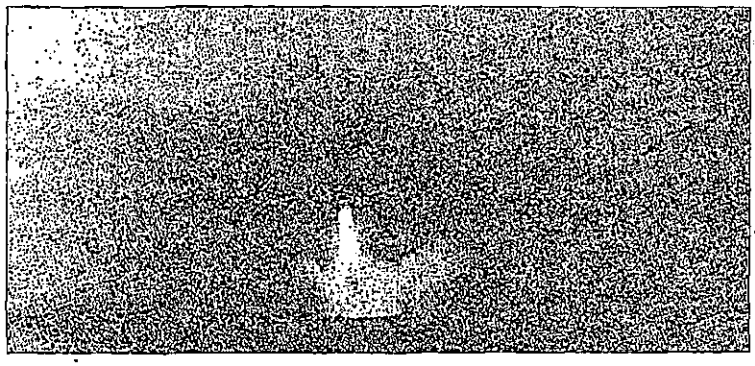
The Final Overflights of the Soviet Union, 1959-1960

THE U-2 AND THE "MISSILE-GAP" DEBATE

Despite President Eisenhower's reluctance to send U-2s over the Soviet Bloc, he once again authorized overflights in the summer of 1959, after a pause of more than a year. The overriding factor in his decision was the growing "missile-gap" controversy, which had its roots in a series of dramatic Soviet announcements during the second half of 1957. The first announcement revealed the successful test of an intercontinental ballistic missile in August. Then in October, the Soviets announced the successful orbiting of the world's first artificial earth satellite, Sputnik. One month later the Soviets orbited a second satellite containing a dog and a television camera. To many Americans, including some influential members of Congress, the Soviet Union's space successes seemed to indicate that its missile program was ahead of that of the United States. By the spring of 1958, after the United States had successfully launched several satellites, fears of a space technology gap between the two superpowers had eased. By the end of the year, however, new concerns arose that the Soviet Union was producing a missile arsenal that would be much larger than that of the United States. This was the famous missile gap that received widespread publicity beginning in early 1959.[1]

The missile-gap controversy was fueled by Soviet boasts about the success of their missile program. On 4 December 1958, a Soviet delegate to the Geneva Conference on Surprise Attack stated: "Soviet ICBMs are at present in mass production." Five days later, Soviet

[1] For an overview of the controversy, see Roy E. Licklider, "The Missile Gap Controversy," *Political Science Quarterly* 85 (1970):600-615.



Launch of Sputnik, 4 October 1957

Premier Nikita Khrushchev asserted that the Soviet Union had an ICBM capable of carrying a 5-megaton nuclear warhead 8,000 miles. These statements seemed all the more ominous because, during this same month of December, the first attempt to launch the new US Titan ICBM failed. In reality, all of the Soviet statements were sheer propaganda; they had encountered difficulties with the SS-6 ICBM, and the program was at a standstill. As a result, there were no ICBM launches from Tyuratam between 29 May 1958 and 17 February 1959, a space of almost nine months.[2]

To conceal the difficulties in their missile program, Soviet leaders continued to praise its alleged successes. At the beginning of February 1959, Khrushchev opened the Soviet Communist Party Congress in Moscow by claiming that "serial production of intercontinental ballistic rockets has been organized." Several months later Soviet Defense Minister Rodion Malinovsky stated that these missiles were capable of hitting "precisely any point" and added, "Our army is equipped with a whole series of intercontinental, continental and other rockets of long, medium and short range." When asked at a press conference to comment on Malinovsky's statement, President Eisenhower replied, "They also said that they invented the flying machine and the automobile and the telephone and other things. . . . Why should you be so respectful of this statement this morning, if you are not so respectful of the other three?"[3] Nevertheless, the Soviet statements were taken at face value by most Americans, including many members of the intelligence community.

---

[2] Lawrence Freedman, *US Intelligence and the Soviet Strategic Threat*, 2nd. ed. (Princeton: Princeton University Press, 1986), pp. 69-70.

[3] Ford Eastman, "Defense Officials Concede Missile Lag," *Aviation Week*, 9 February 1959, pp. 26-27.

As concern about Soviet missile progress increased, even the interruption in Soviet ICBM testing was seen as evidence of a Soviet advantage. Although the CIA correctly reasoned that the Soviets were experiencing difficulties in developing an operational ICBM, the Air Force assumed that the Soviets had halted testing because the missile was ready for deployment.[4]

The controversy intensified early in February 1959, when Secretary of Defense Neil H. McElroy testified before the Senate Preparedness Investigating Committee on Soviet missile capabilities for the next few years. McElroy told the Senators that in the early 1960s the Soviet Union might have a 3 to 1 advantage over the United States in operational ICBMs. McElroy stressed that the gap would be temporary and that at its end the United States would enjoy a technological advantage because it was concentrating on developing the more advanced solid-fueled missiles rather than increasing the number of obsolescent liquid-fueled missiles, but it was his mention of a 3 to 1 missile gap that made the headlines. Administration critics such as Senator Stuart Symington quickly charged that the actual gap would eventually be even larger.[5]

Faced with rising public and Congressional concern about the missile gap, Defense Department officials pressed President Eisenhower to authorize renewed overflights to gather up-to-date information about the status of the Soviet missile program. Following a National Security Council meeting on 12 February, Chairman of the Joint Chiefs of Staff Twining, Secretary of Defense McElroy, and Deputy Secretary of Defense Quarles stayed behind to talk to the President about overflights. They hoped that the need to refute criticism of the missile gap from Symington and other Democratic Senators would persuade the President to loosen his policy on the use of the U-2. McElroy pointed out that no matter how often Allen Dulles briefed these critics, they would not believe his reassurances about the absence of a missile gap without positive proof such as photographs. More overflights would be needed to obtain the kinds of photographs required.

The President was not swayed by these arguments. Noting that the reconnaissance satellite project was "coming along nicely," he stated that U-2 flights should be "held to a minimum pending the

[4] Freedman, *US Intelligence*, p. 70.

[5] "What About the Missile Gap?" *Time*, 9 February 1959, pp. 11-13.

availability of this new equipment." Quarles objected that the satellites would not be ready for up to two years, but the President replied that this did not matter because the Soviets would not be able to build a first-strike force of ICBMs in the near future. President Eisenhower finally conceded that "one or two flights might possibly be permissible," but he ruled out "an extensive program." In light of the "crisis which is impending over Berlin" he did not want to be provocative.[6]

As the missile-gap controversy raged, President Eisenhower stuck to his refusal to permit overflights of the Soviet Union, although the Soviet Union's resumption of ICBM testing almost persuaded him to change his mind. On 10 April 1959, the President tentatively approved several overflights, but, on the following day, he called in McElroy and Bissell to inform them that he was withdrawing his authorization, explaining that "there seems no hope for the future unless we can make some progress in negotiation." Eisenhower remained worried by "the terrible propaganda impact that would be occasioned if a reconnaissance plane were to fail." Although he agreed that new information was necessary, especially in light of the "distortions several senators are making of our military position relative to the Soviets," Eisenhower believed that such information would not be worth "the political costs."[7]

The President remained willing to consider flights that did not overfly Soviet territory, and in June he authorized two electronic intelligence collection missions along the Soviet-Iranian border. The two missions of Operation HOT SHOP took place on 9 and 18 June 1959. The first of these missions was noteworthy because it involved both an Agency U-2 and an Air Force RB-57D Canberra. The two aircraft cruised along the Soviet border and made the first telemetry intercept ever from a Soviet ICBM during first-stage flight, 80 seconds after launch.[8]

Efforts to persuade the President to authorize penetration missions continued. On 7 July 1959, Allen Dulles and Richard Bissell met with Eisenhower to discuss the possibility of a penetration flight

---

[6] Andrew J. Goodpaster, Memorandum for the Record, 12 February 1959, WHOSS Alpha, DDEL (~~TS~~); Ambrose, *Eisenhower: The President*, pp. 513-514; Beschloss, *Mayday*, p. 173.

[7] Quoted in Ambrose, *Eisenhower: The President*, pp. 514-515; Beschloss, *Mayday*, p. 176.

[8] Mission folders 4120 (9 June 1959) and 4121 (18 June 1959), OSA records, job 67-B-972, boxes 10 and 11 (~~TS~~ Codeword).

to gather intelligence on the Soviet missile program. Discussions continued the following day with the addition of Secretary of State Herter, who stated in support of the CIA proposal that "the intelligence objective outweighs the danger of getting trapped." The strong backing of the proposed overflight by both CIA and the State Department finally convinced President Eisenhower to approve the mission.[9]

On 9 July 1959, more than 16 months after the previous overflight of the Soviet Union, a U-2 equipped with a B camera left Peshawar, Pakistan, flew over the Urals, and then crossed the missile test range at Tyuratam. This mission, known as Operation TOUCHDOWN, produced excellent results. Its photography revealed that the Soviets were expanding the launch facilities at Tyuratam. While this overflight was under way, another U-2 flew a diversionary mission along the Soviet-Iranian border.[10]

Despite its success, this overflight remained an isolated incident. President Eisenhower was unwilling to authorize additional overflights of the Soviet Union, in part because he did not wish to increase tension before Premier Khrushchev's visit to the United States scheduled for 15-27 September 1959. Nevertheless, the President still wanted as much intelligence on the Soviet missile program as possible. Because the Soviets were conducting an extensive program of missile tests in mid-1959, Eisenhower authorized a steady stream of the less provocative electronic intelligence (ELINT)-gathering missions (14 in all) along the Soviet border during the remainder of the year.[11]

Within the United States, concern about the Soviet missile program continued to grow. On 12 September 1959 the Soviets scored another space success when their Luna 2 rocket reached the moon, and Khrushchev stressed this success when he arrived in the United States three days later. He also boasted of Soviet missile progress in private conversations with President Eisenhower, while making no

---

[9] Andrew J. Goodpaster, Memorandum for the Record, 7 July 1959 (~~TS~~); idem, Memorandum of Conference with the President, 8 July 1959, WHOSS, Alpha, DDEL ~~(FPS)~~.

[10] Mission folder 4125 (9 July 1959), OSA records (~~TS~~ Codeword).

[11] *OSA History*, chap. 19, annex 120, pp. 12-14 (~~TS~~ Codeword).

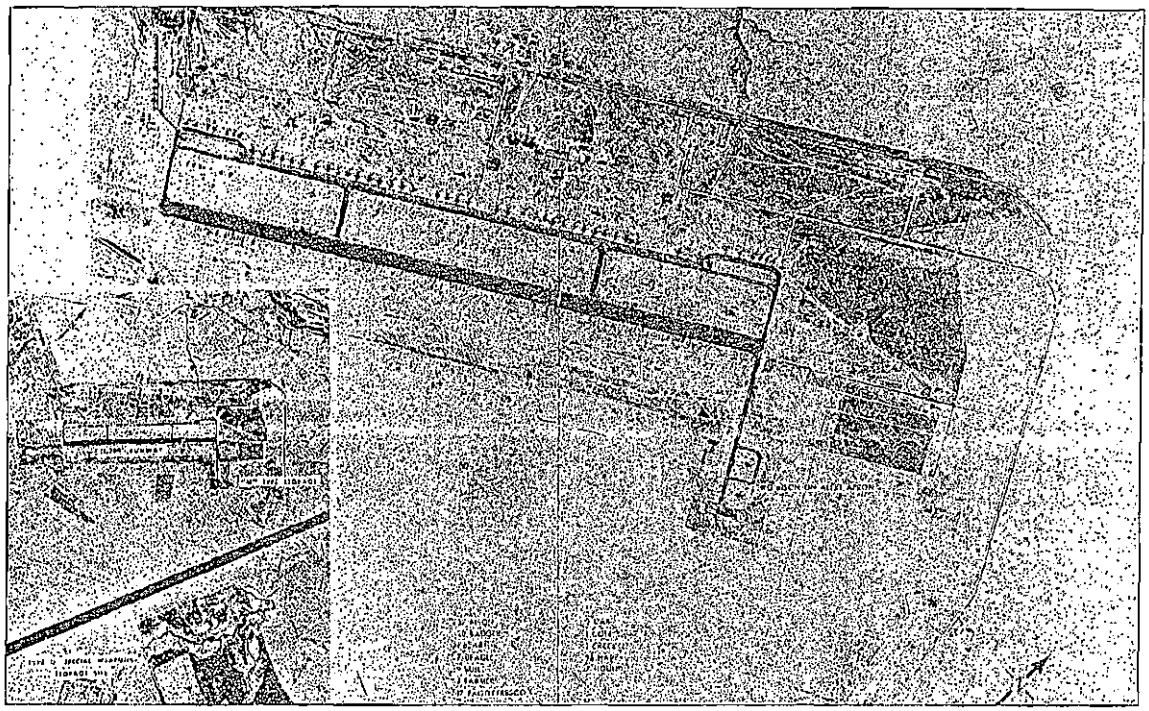
mention of overflights by the United States. After the trip was over, Khrushchev and other leading Soviet officials continued to make exaggerated claims about the extent of their missile force, adding to the confusion and concern within the US intelligence community. Thus in November 1959, Soviet Premier Khrushchev told a conference of journalists, "Now we have such a stock of rockets, such an amount of atomic and hydrogen weapons, that if they attack us, we could wipe our potential enemies off the face of the earth." He then added that "in one year, 250 rockets with hydrogen warheads came off the assembly line in the factory we visited." [12] Because the Soviet Union had been launching at least one missile per week since early fall, US policymakers placed great weight on his remarks.

Despite the intelligence community's intense interest in the Soviet Union's nuclear and missile programs, President Eisenhower did not authorize any more overflights of the Soviet Union during the remainder of the year. On the other hand, he raised no objections to (and probably welcomed) the first British overflight of the Soviet Union in December 1959. For almost a year, the RAF pilots of Detachment B had been ready to fly over the Soviet Union, but Prime Minister Harold Macmillan had not previously authorized any such missions because of his own visit to the Soviet Union, several international meetings, and other state visits. As a result, British U-2 missions had been confined to the Middle East. Now that the Prime Minister's approval had been obtained, Detachment B conducted Operation HIGH WIRE with an RAF pilot. Squadron leader Robert Robinson left Peshawar on 6 December and overflew Kuybyshev, Saratov Engels Airfield, and the Kapustin Yar Missile Test Range before landing at Adana. The mission photography was excellent, but it did not provide intelligence on Soviet ICBMs, which were tested at Tyuratam, not Kapustin Yar.[13]

Because there had been so few overflights in 1958 and 1959, many questions about the Soviet missile program remained unanswered. Within the intelligence community there was still considerable disagreement over the size of the Soviet missile force. Thus, during testimony before the US Senate in January 1960, DCI Allen Dulles, Secretary of Defense Thomas Gates, and Air Force Chief of

[12] William E. Burrows, *Deep Black: Space Espionage and National Security* (New York: Random House, 1987), p. 101.

[13] *OSA History*, chap. 11, p. 33 (~~CPS~~ Codeword); Mission folder 8005 (6 December 1959), OSA records, job 67-B-972, box 13 (~~CPS~~ Codeword).

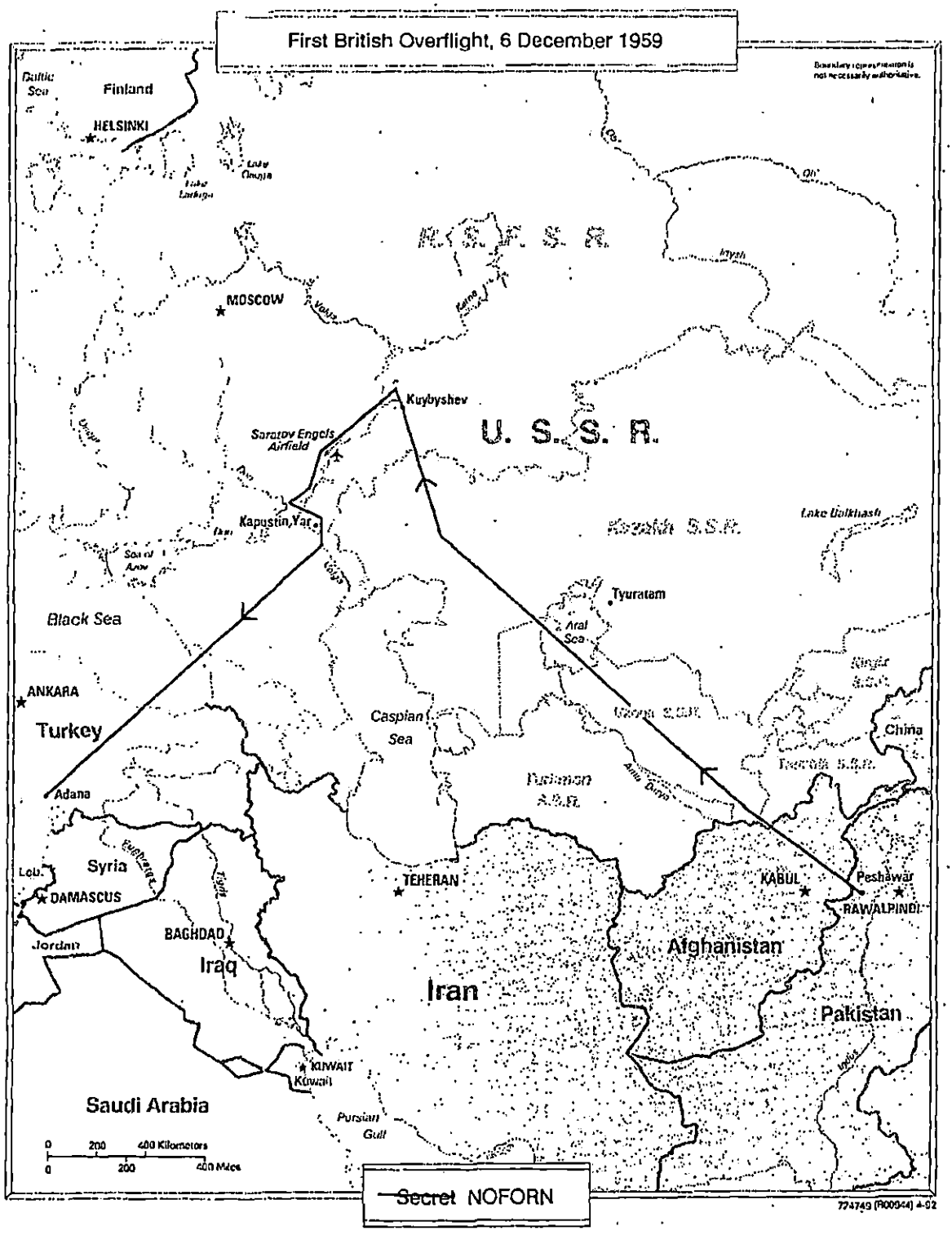


Saratov Engels Airfield, 6 December 1959

Staff Nathan Twining each gave different figures for the number of deployed Soviet missiles. Although the CIA figures were based on evidence gained from overflights, Dulles could not reveal this fact to the Senate and, therefore, faced very sharp questioning.[14]

As a result of these Senate hearings, Dulles was determined to obtain permission for more overflights in order to settle the missile-gap question once and for all and end the debate within the intelligence community. To accomplish this, Dulles proposed photographing the most likely areas for the deployment of Soviet missiles. At this time there was still no evidence of SS-6 ICBM deployment outside the Tyuratam missile test range. Because the SS-6 was extremely large and liquid fueled, analysts believed these missiles could only be deployed near railroads. Existing U-2 photography showed railroad tracks going right to the launching pad at the test site. Dulles, therefore argued that SS-6 installations could easily be located by flying along railroad lines. Dulles was supported by members of the

[14] Licklider, "Missile Gap Controversy," pp. 608-609.


First British Overflight, 6 December 1959
Baltic Sea
Finland
HELSINKI
Lake Onega
Lake Ladoga
R. S. F. S. R.
MOSCOW
Volga
Kuybyshev
Saratov Engels Airfield
U. S. S. R.
Kapustin Yar
Kazakh S.S.R.
Lake Balkhash
Sea of Azov
Black Sea
Tyuratam
Aral Sea
ANKARA
Turkey
Caspian Sea
Uzbek S.S.R.
Tajik S.S.R.
China
Adana
Turkmen S.S.R.
Amu Darya
Syria
Euphrates
Leb.
DAMASCUS
Tigris
TEHERAN
KABUL
Peshawar
RAWALPINDI
Jordan
BAGHDAD
Iraq
Afghanistan
Iran
Pakistan
Indus
KUWAIT
Kuwait
Saudi Arabia
Persian Gulf
0 200 400 Kilometers
0 200 400 Miles
~~Secret~~ NOFORN
774749 (R00944) 4-92

President's Board of Consultants on Foreign Intelligence Activities. At a meeting of the board on 2 February 1960, Gen. James Doolittle urged President Eisenhower to use overflights of the Soviet Union to the maximum degree possible. The President's response, as summarized in General Goodpaster's notes of the meeting, showed that the upcoming summit meeting was already an important factor in his attitude toward U-2 flights: "The President said that he has one tremendous asset in a summit meeting, as regards effect in the free world. That is his reputation for honesty. If one of these aircraft were lost when we are engaged in apparently sincere deliberations, it could be put on display in Moscow and ruin the President's effectiveness." [15]

A few days later, another U-2 took to the sky on a mission over the Soviet Union. As in December, the pilot was British, and the mission had been ordered by Prime Minister Macmillan. On 5 February 1960, a Detachment B U-2C with squadron leader John MacArthur at the controls left Peshawar, Pakistan, to conduct Operation KNIFE EDGE. The plane overflew the Tyuratam Missile Test Range, headed northwest to Kazan', and then turned south, photographing long stretches of the Soviet rail network. The excellent photography from this mission did not reveal a single missile site, but analysts did discover a new Soviet bomber, dubbed the BACKFIN, at Kazan'.[16]

Despite the outcome of this mission, the missile-gap debate continued. The Air Force still insisted that the Soviets had deployed as many as 100 missiles. The Army, Navy, and CIA, however, doubted that any had been deployed, because none could be found. Additional U-2 photography was needed to settle the debate. In mid-February, President Eisenhower reviewed plans for four additional U-2 missions. The success of the two British missions, along with the absence of Soviet protests, made the President more willing to consider a resumption of US overflights, and he agreed to allow one mission to be flown during the month of March. The President's continued restrictions upon the use of the U-2 disturbed DCI Dulles, who sent a memorandum to the National Security Council on 1 March 1960 asserting that the cardinal objective of obtaining information on Soviet missile deployment could be better achieved if the U-2 were given freer rein.[17]

---

[15] Ambrose, *Eisenhower: The President*, p. 568; Beschloss, *Mayday*, p. 233.

[16] Mission folder 8009 (5 February 1960), OSA records, job 67-B-972, box 13 (TS Codeword); *OSA Chronology*, p. 25 (TS Codeword).

[17] *OSA Chronology*, p. 25 (TS Codeword); Philip K. Edwards, "The President's Board: 1956-60," *Studies in Intelligence* 13 (Summer 1969):118 (S).

In authorizing another overflight of the Soviet Union, President Eisenhower directed that it be conducted before 30 March. Because of complications in getting permission from Pakistan to use the airfield at Peshawar, however, the mission could not be staged in March, and the President agreed to extend his deadline until 10 April 1960. One day before the expiration of this deadline, a U-2 equipped with a B-camera took off from Peshawar on the last successful overflight of the Soviet Union, Operation SQUARE DEAL. As had been the case during the previous two overflights, a second U-2 flew a diversionary mission along the Soviet-Iranian border. After leaving Peshawar, mission 4155 headed first for Saryshagan, where it obtained the first pictures of two new Soviet radars, the HEN HOUSE and HEN ROOST installations. The U-2 then flew to the nuclear testing site at Semipalatinsk. Returning to the Saryshagan area, it crisscrossed the railroad network there and then proceeded to Tyuratam, where it photographed a new two-pad, road-served launch area that suggested a new Soviet missile was in the offing.[18]

In his memoirs Nikita Khrushchev remarked that this U-2 should have been shot down, "but our antiaircraft batteries were caught napping and didn't open fire soon enough." Khrushchev explained that Soviet missile designers had developed a high-altitude antiaircraft missile and batteries of this missile had been deployed near known targets of the U-2.[19]

The CIA already had strong indications of improvements in the Soviet air defense system, and early in 1960 the Development Projects Division had asked Air Force experts at the Air Technical Intelligence Center (ATIC) for a frank assessment of Soviet capabilities against the U-2. On 14 March 1960, Col. William Burke, acting chief of the DPD, relayed the ATIC assessment to Richard Bissell:

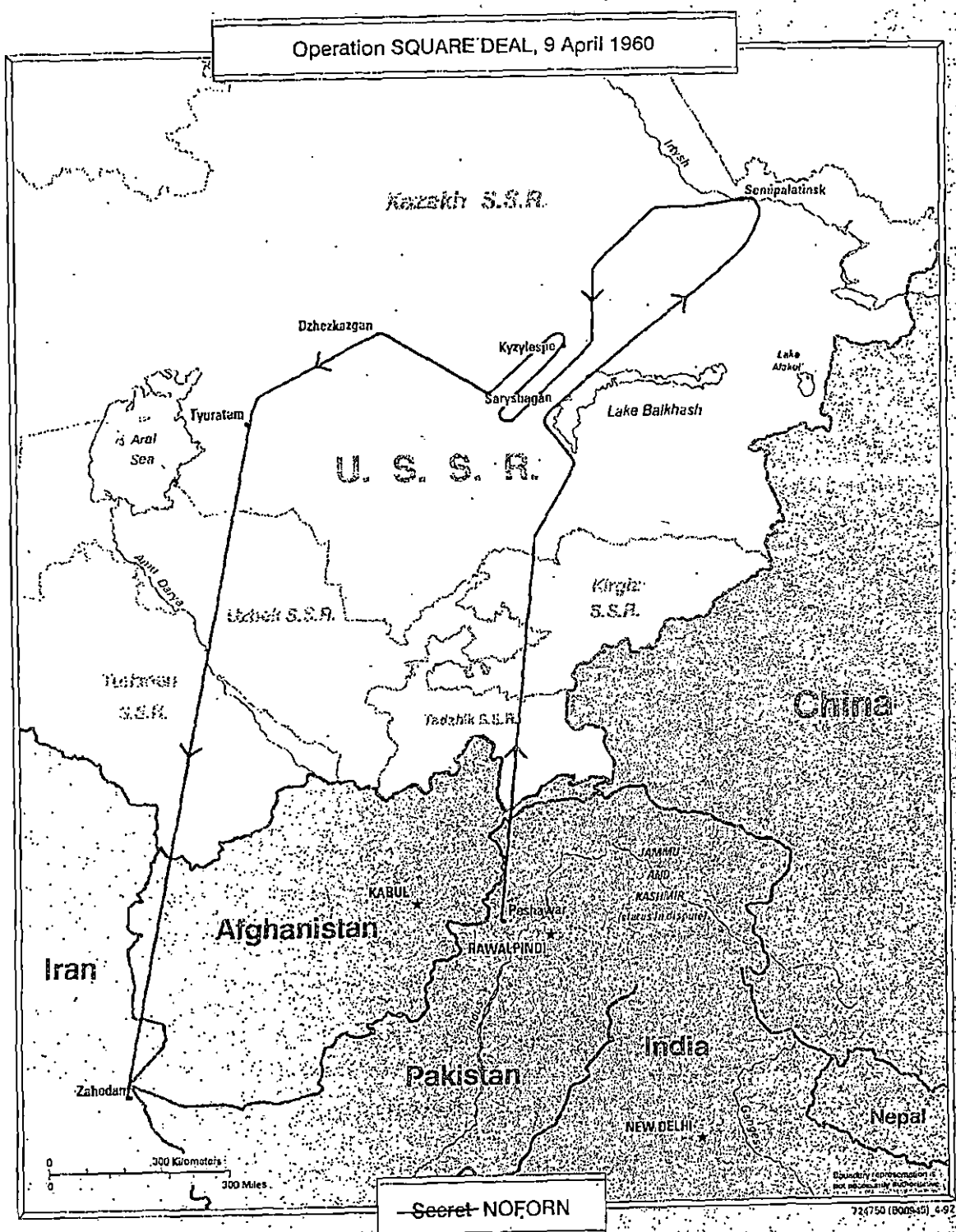
> *The greatest threat to the U-2 is the Soviet SAM. Although the ATIC analysis concedes a remote possibility that the SAM may be less effective than estimated, their present evaluation is that the SAM (Guideline) has a high probability of successful intercept at 70,000 feet providing that detection is made in sufficient time to alert the site.*[20]

---

[18] Mission folder 4155, 9 April 1960, OSA records, job 67-B-328, box 6 (~~TS~~ Codeword).

[19] Nikita S. Khrushchev, *Khrushchev Remembers: The Last Testament* (Boston: Little, Brown, & Co., 1974), pp. 443-444.

[20] Memorandum for Richard M. Bissell, Deputy Director (Plans), from Col. William Burke, Acting Chief, DPD, "Evaluation of Proposed CHALICE Operations," 14 March 1960, IC Staff, COMIREX records, job 33-T-123A, box 10, "CHALICE (General)" (~~TS~~ Codeword).


Operation SQUARE DEAL, 9 April 1960
Kazakh S.S.R.
Irtysh
Semipalatinsk
Dzhezkazgan
Kyzylespe
Sarysbagan
Lake Alakol'
Lake Balkhash
Tyuratam
Aral Sea
U. S. S. R.
Amu Darya
Uzbek S.S.R.
Kirgiz S.S.R.
Turkmen S.S.R.
Tadzhik S.S.R.
China
JAMMU AND KASHMIR (status in dispute)
KABUL
Peshawar
RAWALPINDI
Afghanistan
Iran
India
Pakistan
Zahodan
NEW DELHI
Nepal
0 300 Kilometers
0 300 Miles
~~Secret~~ NOFORN
72(750 (B00945) 4-92

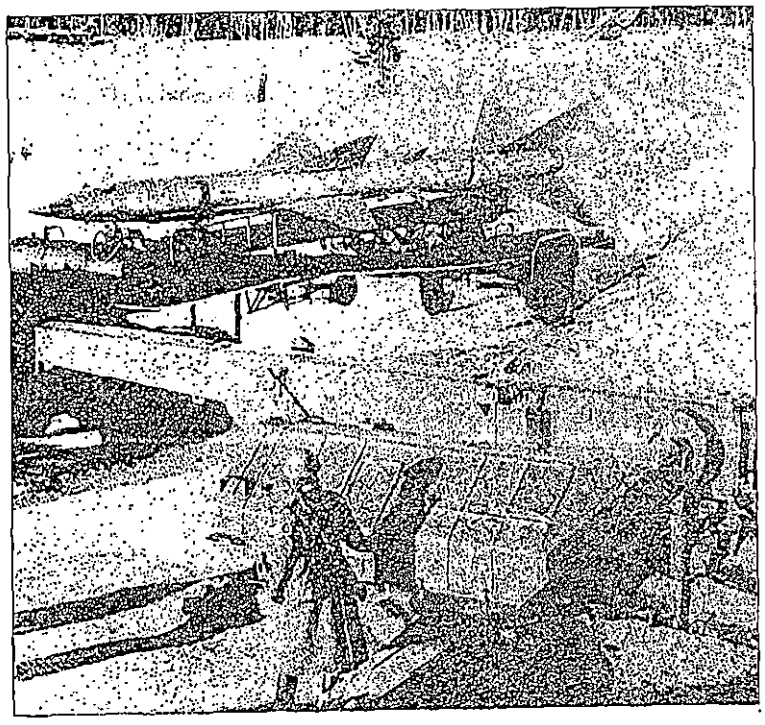
One of the reasons why Operation SQUARE DEAL had been selected for the 9 April flight was that mission planners believed that penetration from the Pakistan/Afghanistan area offered the greatest chance of escaping detection by the Soviet air defense system. Colonel Burke's 14 March letter recommending SQUARE DEAL as the preferred route for the next overflight had stated, "There is a reasonable chance of completing this operation without detection." Escaping detection had become important because, if the Soviet SAMs received sufficient advanced warning, they posed a major threat to the U-2.

CIA hopes that flights from Pakistan or Afghanistan might go undetected proved false. On the 9 April overflight, the U-2's ELINT-collection unit (System VI) indicated Soviet tracking at a very early stage of the mission. Although the Soviets failed to intercept the U-2, their success at tracking it should have served as a warning against future overflights from Pakistan (or anywhere else, for that matter). On 26 April 1960, Colonel Burke informed Richard Bissell that "experience gained as a result of Operation SQUARE DEAL indicates that penetration without detection from the Pakistan/Afghanistan area may not be as easy in the future as heretofor." [21] Unfortunately, neither Colonel Burke nor Richard Bissell took the logical step of recommending the cessation of overflights now that the risks had increased substantially. The lure of the prospective intelligence gain from each mission was too strong, and the Soviets' lack of success at interception to date had probably made the project staff overconfident. Furthermore, both DCI Allen Dulles and the President's Board of Consultants on Foreign Intelligence Activities were pressing for more photos of the Soviet Union in order to settle the missile-gap debate raging in the intelligence community and Congress.

THE LAST OVERFLIGHT: OPERATION GRAND SLAM

Even before the 9 April overflight took place, President Eisenhower had consented on 28 March to an additional overflight during the month of April. His willingness to allow yet another overflight was

[21] Memorandum for Richard M. Bissell, Deputy Director (Plans), from Colonel Burke, Acting Chief, DPD, "Operational Priority of Proposed CHALICE Missions," 26 April 1960, IC Staff, COMIREX records, job 33-T-123A, box 10, "CHALICE (General)" (~~TS~~ Codeword).



SA-2 surface-to-air missile

strengthened when the Soviet Union did not protest the 9 April mission. As Presidential science adviser George Kistiakowsky later remarked about the lack of protest, "This was virtually inviting us to repeat the sortie."[22]

Although President Eisenhower had authorized another overflight for April, he left the designation of its targets up to the experts at the CIA. Of the three missions that remained under consideration, one—Operation SUN SPOT—would overfly southern targets, Tyuratam and Vladimirovka, while the other two would cover railroad networks in the north-central portion of the Soviet Union. The intelligence community had been interested in this area ever since late 1959, when there were indications that the Soviets were building an SS-6 launch facility there. This was the first indication that SS-6s might be located anywhere other than Tyuratam testing facility, where the missiles were launched from a general purpose launching pad. The intelligence community was anxious to obtain photography of a

[22] George B. Kistiakowsky, *A Scientist at the White House* (Cambridge: Harvard University Press, 1976), p. 328.

deployed SS-6 site because it could provide exemplars for photointerpreters to use in searching subsequent overhead photography for similar installations.[23]

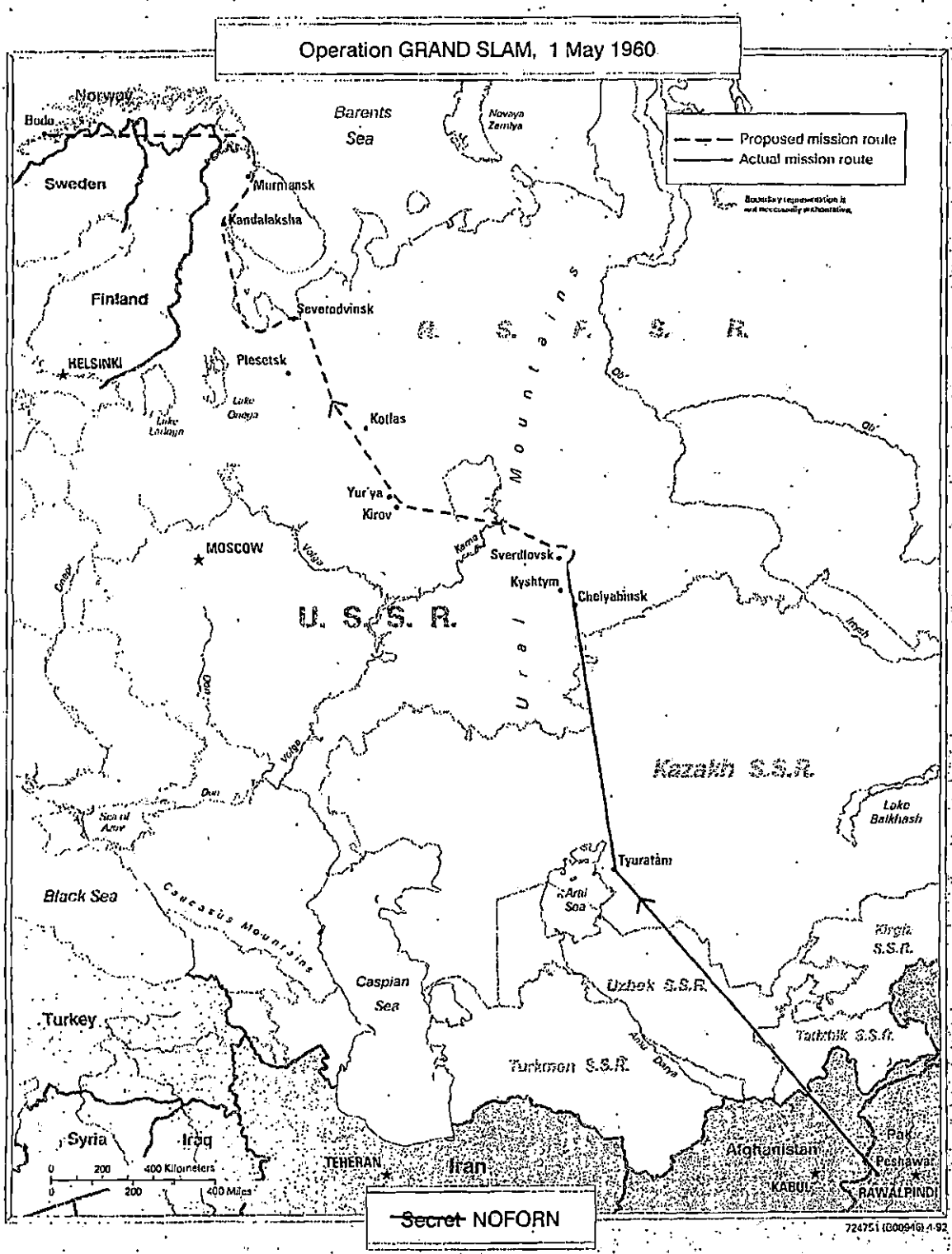
The two proposed overflights that would cover the northern railroad lines received the strongest consideration. Both plans contained new features. Operation TIME STEP called for a U-2 to take off from the USAF base at Thule, Greenland, which would be the first overflight staged from this base. The aircraft would then fly over Novaya Zemlya on its way to cover the railroad lines from the Polyarnyy Ural Mountains to Kotlas. The return flight would be over Murmansk with the landing to take place at either Bodo or Andoya on Norway's northeast coast. The other proposed overflight, Operation GRAND SLAM, was the first U-2 mission planned to transit the Soviet Union; all previous missions had penetrated not more than halfway and then left in the general direction from which they came. GRAND SLAM proposed to fly across the Soviet Union from south to north, departing from Peshawar, Pakistan, and landing at Bodo, Norway. The mission would overfly Tyuratam, Sverdlovsk, Kirov, Kotlas, Severodvinsk, and Murmansk.

The two preferred missions both required the use of the airfield at Bodo, which had been authorized by senior Norwegian intelligence and military officers. Because the Bodo airfield was involved in NATO maneuvers taking place in the Barents Sea area, Bissell informed the White House that neither mission could be flown before 19 April. Once the maneuvers ended, bad weather over the Soviet Union kept the mission from taking place when it was originally scheduled. Richard Bissell, therefore, asked President Eisenhower for more time, and, on 25 April, General Goodpaster relayed the President's instructions to Bissell that "one additional operation may be undertaken, provided it is carried out prior to May 1. No operation is to be carried out after May 1."[24] The President did not want to fly missions any later than that because the Paris Summit was scheduled to begin on 16 May 1960.

By this time, CIA planners were concentrating on Operation GRAND SLAM as the most likely route for the proposed mission because it offered the best chance of photographing suspected locations

[23] *OSA History*, chap. 12, pp. 35-36 (~~TS~~ Codeword).

[24] Ambrose, *Eisenhower: The President*, p. 569; Beschloss, *Mayday*, p. 10.


Operation GRAND SLAM, 1 May 1960
Proposed mission route
Actual mission route
Norway
Bodo
Sweden
Finland
HELSINKI
Barents Sea
Novaya Zemlya
Murmansk
Kandalaksha
Severodvinsk
Plesetsk
Lake Onega
Lake Ladoga
Kotlas
Yur'ya
Kirov
MOSCOW
Volga
Kama
Sverdlovsk
Kyshtym
Chelyabinsk
U. S. S. R.
Ural Mountains
Ob'
Irtysh
Kazakh S.S.R.
Lake Balkhash
Tyuratam
Aral Sea
Black Sea
Caucasus Mountains
Caspian Sea
Kirgiz S.S.R.
Uzbek S.S.R.
Tadzhik S.S.R.
Turkmen S.S.R.
Amu Darya
Turkey
Syria
Iraq
TEHERAN
Iran
Afghanistan
KABUL
Peshawar
RAWALPINDI
0 200 400 Kilometers
0 200 400 Miles
~~Secret~~ NOFORN
724751 (00094) 1-92

of Soviet ICBM sites. The other proposed overflight, Operation TIME STEP out of Greenland, was more likely to run into bad weather (which would affect both navigation and photography) because the flightpath would remain above 60° north latitude during the entire mission. Furthermore, mission planners opposed this route because of its greater risk. In his letter to Richard Bissell on 14 March 1960, Colonel Burke stated:

> *Operation "TIME STEP" is our last choice because we can assume, with a 90 percent probability of being correct, that we will be detected on entry, tracked accurately throughout the period in denied territory (approximately four hours), and will evoke a strong PVO [Soviet Air Defense] reaction. This flight plan would permit alerting of SAM sites, and pre-positioning of missile equipped fighters in the Murmansk area (point of exit) thus enhancing the possibility of successful intercept. In addition, we must assume that even were the Soviets unable to physically interfere with such an incursion, sufficient evidence will be available to permit them to document a diplomatic protest should they desire to do so.*[25]

The concerns raised by Colonel Burke about TIME STEP should also have been raised about Operation GRAND SLAM, which would be the most adventuresome overflight to date because it proposed covering so much of the Soviet Union. If the Soviets could track the U-2 early in the mission, they would have plenty of time to prepare to intercept the aircraft.

The pilot selected for Operation GRAND SLAM was Francis Gary Powers, the most experienced U-2 pilot in the program. Powers had joined the project in May 1956 and had flown 27 operational missions in the U-2, including one each over the Soviet Union and China as well as six along the Soviet border.

To prevent the U-2 from being seen at Peshawar, project managers decided to ferry the aircraft from Adana to Pakistan the night before the scheduled flight. Once the plane was refueled and its camera was loaded, it would take off at daybreak, with little if any exposure to local residents because of darkness and its short stay—less than six

[25] Memorandum for Richard M. Bissell, Deputy Director (Plans), from Col. William Burke, Acting Chief, DPD, "Evaluation of Proposed CHALICE Operations," 14 March 1960, IC Staff, COMIREX records, job 33-T-123A, box 10, "CHALICE (General)" (~~TS~~ Codeword).

hours on the ground. Originally scheduled for Thursday, 28 April, GRAND SLAM was canceled because of bad weather over the northern Soviet Union. This had been the case for the past several weeks. When this flight was canceled, the U-2 returned to Adana before sunrise. That evening the U-2 flew back to Peshawar for another attempt to stage the mission early on the 29th, but bad weather again forced cancellation of the mission, and the U-2 returned to Adana. Because of continued bad weather over the target areas, no mission was planned for Saturday, 30 April.[26]

Meanwhile, the plane ferried to Peshawar on 27 and 28 April had accumulated so many hours of flight time that it had to be removed from service for periodic maintenance. A different aircraft was, therefore, ferried to Peshawar on Saturday night, 30 April. This aircraft, article 360, had made a crash landing in Japan during the previous September (see chapter 5). Although it had been refurbished by Lockheed and now had the more powerful J75 engine that would give it greater altitude, pilots did not completely trust this aircraft and considered it a "hangar queen." As Powers noted in his memoirs, "Its current idiosyncrasy was one of the fuel tanks, which wouldn't feed all its fuel."[27] The aircraft was equipped with a B-model camera, a System-VI electronic intelligence unit, and a System-IXB device, which generated false-angle information in response to the radar pulses used by some Soviet airborne-missile fire-control systems.

Operation GRAND SLAM, mission 4154 and the 24th deep-penetration overflight of the Soviet Union, began almost 30 minutes late on Sunday, 1 May 1960, a delay due to difficulty in getting takeoff approval from Washington. This delay was caused by communications problems that are typical at sunrise and sunset during spring and autumn, when the ionosphere will not support reliable communications. In attempting to relay the authorization message, the radio operator in Adana was unable to reach the base in Peshawar, whose codename was HBJARGON. Realizing that neither the prearranged nighttime nor daytime frequencies were working, the operator began sending a message in the clear, using one of the guard frequencies in the transition area between the daytime and nighttime frequencies. The radio operators at Peshawar kept hearing the Morse

[26] Mission folder 4154 (1 May 1960), OSA records (TS Codeword).

[27] Powers, *Operation Overflight*, p. 76.

code letters JGOHB, JGOHB as they tuned from one prearranged frequency to the other. Then one of the Peshawar operators decided to tune in the guard frequency where the Morse transmission was strongest. He was able to discern a break in the letters, making the message read "HBJGO HBJGO." The Peshawar operators realized this stood for "HBJARGON Go." The detachment chief, Col. William Shelton, who had been waiting anxiously inside the radio van for a "Go" or "No Go" message, leaped from the van and ran across the field to give the signal for takeoff to Powers, who was sitting in the U-2C at the end of the runway.[28]

Powers started his takeoff roll at 0159Z on 1 May 1960. Once airborne, Powers guided his aircraft toward Afghanistan. Following standard operating procedure, Powers clicked his radio switch when he reached penetration altitude of 66,000 feet, which signaled the operations unit at Peshawar that everything aboard the aircraft was working and the mission would proceed as planned. Aside from this simple signal, Powers and all U-2 pilots maintained strict radio silence during penetration missions.

Powers' first target was the Tyuratam Missile Test Range after which he headed for Chelyabinsk, just south of Sverdlovsk. The planned route would take him over Kyshtym, Sverdlovsk, northwest to Kirov, north over Yur'ya and Plesetsk, then to Severodvinsk, northwest to Kandalaksha, north to Murmansk, and, finally, west to Bodo, Norway.

May Day turned out to be a bad time to overfly the Soviet Union. On this major holiday, there was much less Soviet military air traffic than usual, so Soviet radars could easily identify and track Powers' U-2. In addition, the Soviets responded to the intrusion by ordering a ban on civilian air traffic in a large portion of the Soviet Union. Soviet radar began tracking the U-2 when it was still 15 miles south of the Soviet-Afghan border and continued to do so as the aircraft flew across the Central Asian republics. When Powers reached the Tashkent area, as many as 13 Soviet interceptor aircraft scrambled in an unsuccessful attempt to intercept his plane.

Powers never made it past Sverdlovsk. Four and a half hours into the mission, an SA-2 surface-to-air missile detonated close to and just behind his aircraft and disabled it 70,500 feet above the Sverdlovsk

[28] Richard K. Pero, "Message Received—Unfortunately," *Studies in Intelligence* 27 (Winter 1983):29 (S).

area. The plane began spiraling down toward the ground and Powers looked for a way out. Unable to use the ejection seat because centrifugal force had thrown him against the canopy, he released the canopy and prepared to bail out, waiting to arm the destruction device at the last minute, so that it would not go off while he was still in the plane. When he released his seatbelt, however, he was immediately sucked out of the aircraft and found himself dangling by his oxygen hose, unable to reach the destruction switches. Finally, the hose broke and he flew away from the falling aircraft. After he fell several thousand feet, his parachute opened automatically, and he drifted to earth where he was quickly surrounded by farmers and then by Soviet officials.[29] His aircraft had not been destroyed by the crash, and the Soviets were able to identify much of its equipment when they put it on display 10 days later. Even if Powers had been able to activate the destruction device, however, it would not have destroyed the aircraft. The small explosive charge was only designed to wreck the camera.

How had the Soviets succeeded in downing the U-2? Although some CIA project officials initially wondered if Powers had been flying too low through an error or mechanical malfunction, he maintained that he had been flying at his assigned altitude and had been brought down by a near miss of a Soviet surface-to-air missile. This turned out to be the case, for in March 1963, the US air attache in Moscow learned that the Sverdlovsk SA-2 battery had fired a three-missile salvo that, in addition to disabling Powers' plane, also scored a direct hit on a Soviet fighter aircraft sent aloft to intercept the U-2.[30] Mission planners had not known about this SAM site before the mission because they always laid out flight plans to avoid known SAM sites.

THE AFTERMATH OF THE U-2 DOWNING

The first indication that something was wrong with Powers' mission came even before he was overdue at Bodo, Norway. The CIA Operations Center learned on 1 May at 0330 hours Washington time

[29] Powers, *Operation Overflight*, pp. 82-84; Beschloss, *Mayday*, pp. 26-28; Transcript of Debriefing Tapes of Francis Gary Powers, 13 February 1962, Board of Inquiry on the Conduct of Francis Gary Powers, Operations MUDLARK files, OSA records, job 74-B-605, box 6 (S).

[30] Cunningham interview, 4 October 1983 (TS Codeword); *OSA History*, chap. 14, p. 55 (TS Codeword).

that the Soviets had discontinued radar tracking of the flight's progress two hours earlier (0529Z), southwest of Sverdlovsk. Although there was no word from the Soviet Union concerning the missing U-2, key project personnel assembled in the Agency control center that morning (with the exception of Bissell, who was out of town and did not arrive until 1530) to analyze the latest information and discuss courses of action. They quickly established a new project, known as Operation MUDLARK, to gather and evaluate all available information about the downed U-2.[31]

Bissell and the other project officials did not know whether Powers was dead or if the plane and camera had been destroyed, but they believed that there was no way that a pilot could survive a crash from an altitude above 70,000 feet. They, therefore, decided to stick with the standard cover story for U-2 flights: that they were weather flights staged by the National Aeronautics and Space Administration (NASA)—originally the National Advisory Committee for Aeronautics, renamed in 1958. This cover story had been approved by the President in 1956.

By the end of the day, the Operation MUDLARK officials had prepared a statement based on the standard cover story but modified to fit the available information on Powers' flight and to show Adana as the aircraft's base in order to conceal Pakistan's role in the mission. This revised cover story, along with a mission flight plan consistent with it, was sent to the field commander at Adana, to Air Force Europe headquarters, [redacted] to replace the cover story that had been prepared and distributed in advance of the mission. The first announcement of the new cover story came late on 2 May by the Adana base commander, but it did not appear in print until the following day. On Tuesday, 3 May, NASA released a statement about a high-altitude weather plane that was missing on a flight inside Turkey. The statement had been designed to provide an explanation for the presence of wreckage inside the Soviet Union by noting that "the pilot reported over the emergency frequency that he was experiencing oxygen difficulties."[32] Thus, if the Soviets protested and pointed to wreckage inside their borders, NASA could claim that the pilot had lost consciousness and the aircraft had then flown into the Soviet Union before crashing.

25X1, E.O.13526

[31] Geary interview (~~S~~).

[32] Beschloss, *Mayday*, p. 39.

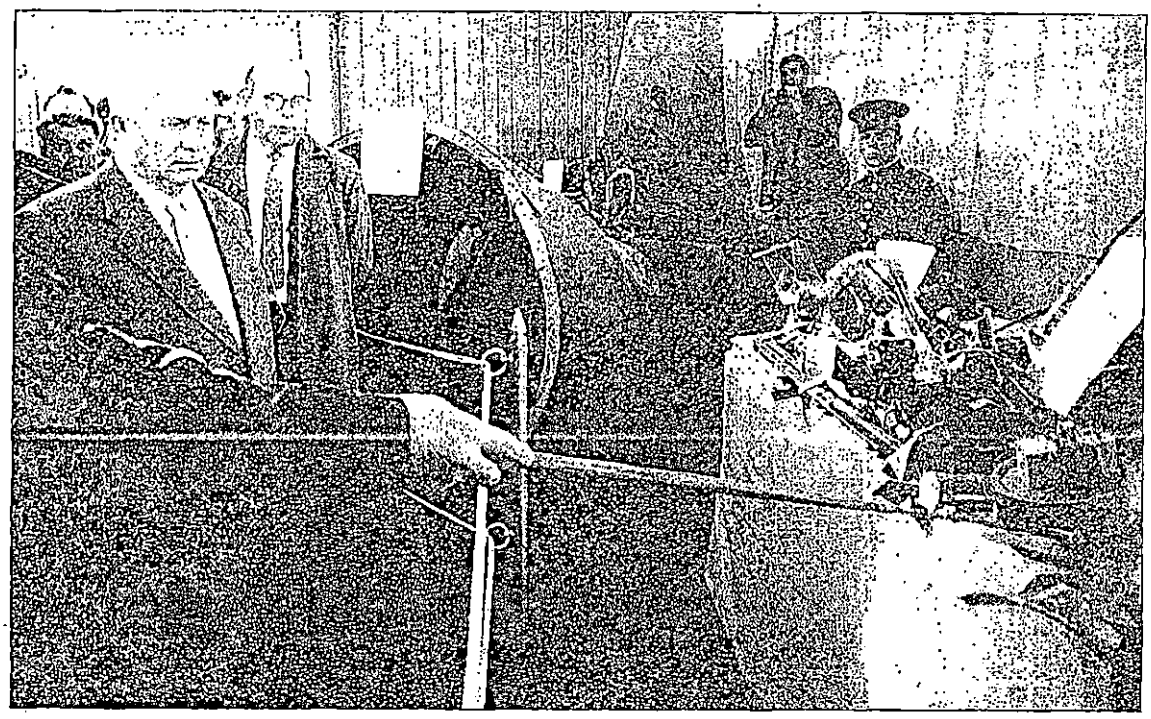
This statement had been prepared for a "best case" scenario, that is to say, one in which neither the pilot nor the plane and film survived. However, pilots had bailed out from extremely high altitudes and survived, and there was even evidence from previous U-2 crashes that much of the aircraft itself could be salvaged. The small destructive charge aboard the U-2 was not sufficient to destroy much more than the camera. The tightly rolled film, which could reveal the exact purpose of the mission even if the pilot and aircraft did not survive, was very hard to destroy. Kelly Johnson later conducted an experiment that revealed film taken out of a completely burned-out aircraft could still provide usable imagery.[33] After almost four years of successful U-2 missions, Richard Bissell and the rest of the Development Projects Division had become overconfident and were not prepared for the "worst case" scenario that actually occurred in May 1960. This failure played directly into the hands of Soviet Premier Nikita Khrushchev, who shrewdly decided to release information about the downed U-2 a little at a time, thereby encouraging the United States to stick with its vulnerable cover story too long. As he later wrote, "Our intention here was to confuse the government circles of the United States. As long as the Americans thought the pilot was dead, they would keep putting out the story that perhaps the plane had accidentally strayed off course and been shot down in the mountains on the Soviet side of the border."[34] The first word from the Soviet Union came on Thursday, 5 May, when Premier Khrushchev announced to a meeting of the Supreme Soviet that a US "spyplane" had been downed near Sverdlovsk. He made no mention of the fate of its pilot.

Khrushchev's announcement aroused considerable interest in the media in the United States, and that same day the State Department and NASA issued another statement that continued the "weather plane" cover story, adding that the pilot became lost during a routine mission near the Caucasus Mountains. Soon afterward, the US Ambassador to Moscow cabled a report to the State Department indicating that the pilot might be alive after all. Two days later, on 7 May 1960, Khrushchev confirmed this report by revealing that the U-2 pilot was alive and had admitted his mission of spying on the Soviet Union.

---

[33] Geary interview.

[34] Khrushchev, *Khrushchev Remembers: The Last Testament*, p. 507.



Khrushchev and the U-2 wreckage

This revelation completely demolished the US cover story, and senior administration officials then debated what the appropriate course of action should be. Allen Dulles offered to take responsibility for the overflight and resign, but President Eisenhower did not want to give the world the impression that he was not in control of his administration. On Wednesday, 11 May, the President read a statement to the press in which he assumed full responsibility for the U-2 mission but left open the question of future overflights, even though four days earlier he had approved the recommendation of his key foreign policy advisers to terminate all provocative intelligence operations against the Soviet Union.[35]

The U-2 affair had its greatest consequences when the long-awaited summit meeting in Paris began less than a week later on 16 May. Soviet Premier Khrushchev insisted on being the first speaker and read a long protest about the overflight, ending with a demand for an apology from President Eisenhower. In his reply

[35] *OSA History*, chap. 14, pp. 14-16 (~~TS~~ Codeword); Beschloss, *Mayday*, pp. 43-66, 243-258.

Eisenhower stated that overflights had been suspended and would not be resumed, but he refused to make a formal apology. At that point the summit ended, as did all hopes for a visit to the Soviet Union by President Eisenhower.

THE WITHDRAWAL OF THE OVERSEAS DETACHMENTS

The loss of Powers' U-2 ultimately resulted in the end of Detachment B in Turkey. As soon as the Development Projects Division learned that Powers was alive in Soviet hands, it immediately evacuated the British pilots from Adana to protect the secret of their involvement in the project. Project officials hoped that flights might eventually resume from Adana, but President Eisenhower's order ending overflights of the Soviet Union made this very unlikely. Less than four weeks later, a coup ousted the government of Turkish Premier Adnan Menderes on the night of 27 May 1960. Because the new government had not been briefed on the U-2, Project Headquarters refused to allow any U-2 flights from Adana, even those necessary for maintaining the aircraft's airworthiness. As a result, no more U-2s flew out of Adana. Instead of being ferried home, three of the four remaining U-2s were disassembled and loaded aboard C-124 cargo planes for the return trip to the United States.[36]

The fourth U-2 remained inside a hangar at Incirlik airbase for several years, looked after by a skeleton crew, in case the Adana installation needed to be reactivated. Finally the decision was made to close down the Adana U-2 facility. During Detachment B's 44 months of active existence, 21 pilots had flown its aircraft, including four RAF pilots and three pilots transferred from the deactivated Detachment A. Fourteen Detachment B pilots were later assigned to other U-2 detachments, but the closing down of Detachment B marked the end of Britain's direct involvement in U-2 operational overflights. A four-man unit of RAF U-2 pilots was stationed at Detachment G, Edwards AFB, until the end of the CIA U-2 program in 1974, but RAF pilots never again conducted an overflight in an Agency U-2.

The loss of Powers' U-2, the resultant failure of the Paris Summit, and the end of U-2 operations in Turkey were just the first in a series of setbacks for the U-2 program. On 8 July 1960, the

[36] *OSA History*, chap. 12, pp. 46-47 (~~TK~~ Codeword).

Japanese Government, faced with growing anti-American sentiment and complaints in the press about the presence of "spyplanes" on Japanese territory, asked the United States to remove the U-2s. The very next day the CIA closed Detachment C; its U-2s were dismantled and returned to the United States aboard C-124s.[37]

In the midst of the furor in Japan, on 1 July 1960, just six weeks after the Paris Summit, Soviet fighter aircraft shot down an Air Force RB-47 on an electronic intelligence collection mission over international waters near the Soviet Union's Kola Peninsula. Two survivors were captured. The Soviet Union claimed that the aircraft had violated its airspace, while the United States denounced the Soviets for downing the plane over international waters. The acrimony exacerbated an already tense international atmosphere.[38]

One additional blow to the U-2 program came in the summer of 1960. NASA, concerned about the damage to its reputation from its involvement in the U-2 affair and hoping to obtain international cooperation for its space program, decided to end its support of the cover story that U-2s were conducting weather research under its auspices.[39]

These developments resulted in a complete halt to all U-2 operations from overseas bases for more than six months. Pilots and aircraft from Detachments B and C were consolidated into Detachment G at Edwards Air Force Base, California, the unit formed after the CIA had vacated the Nevada testing site in 1957 as a result of AEC nuclear testing. Detachment G now comprised eight pilots from Detachment B and three pilots from Detachment C. Because Powers' capture had compromised Project CHALICE, the Agency assigned a new cryptonym to the U-2 effort; henceforth, it was called Project IDEALIST.[40]

---

[37] *OSA Chronology*, p. 28 (~~TS~~ Codeword).

[38] "Mystery of the RB-47," *Newsweek*, 25 July 1960, pp. 36-37; "Nikita and the RB-47," *Time*, 25 July 1960, pp. 30-31.

[39] At a meeting of high-level CIA, NASA, and State Department officials on 31 May 1960, NASA was willing to continue its association with U-2 flights for the time being, but the Administrator of NASA, Dr. Keith Glennan, believed that his agency "would be well advised to disengage from the U-2 program as rapidly as possible." James A. Cunningham, Memorandum for the Record, "Telephone Conversation with Dr. Hugh Dryden, Deputy Director, NASA," 1 June 1960, DPD chrono file #4553-60, OSA records ~~(S)~~.

[40] *OSA History*, chap. 12, pp. 47-49; chap. 16, p. 10 (~~TS~~ Codeword).

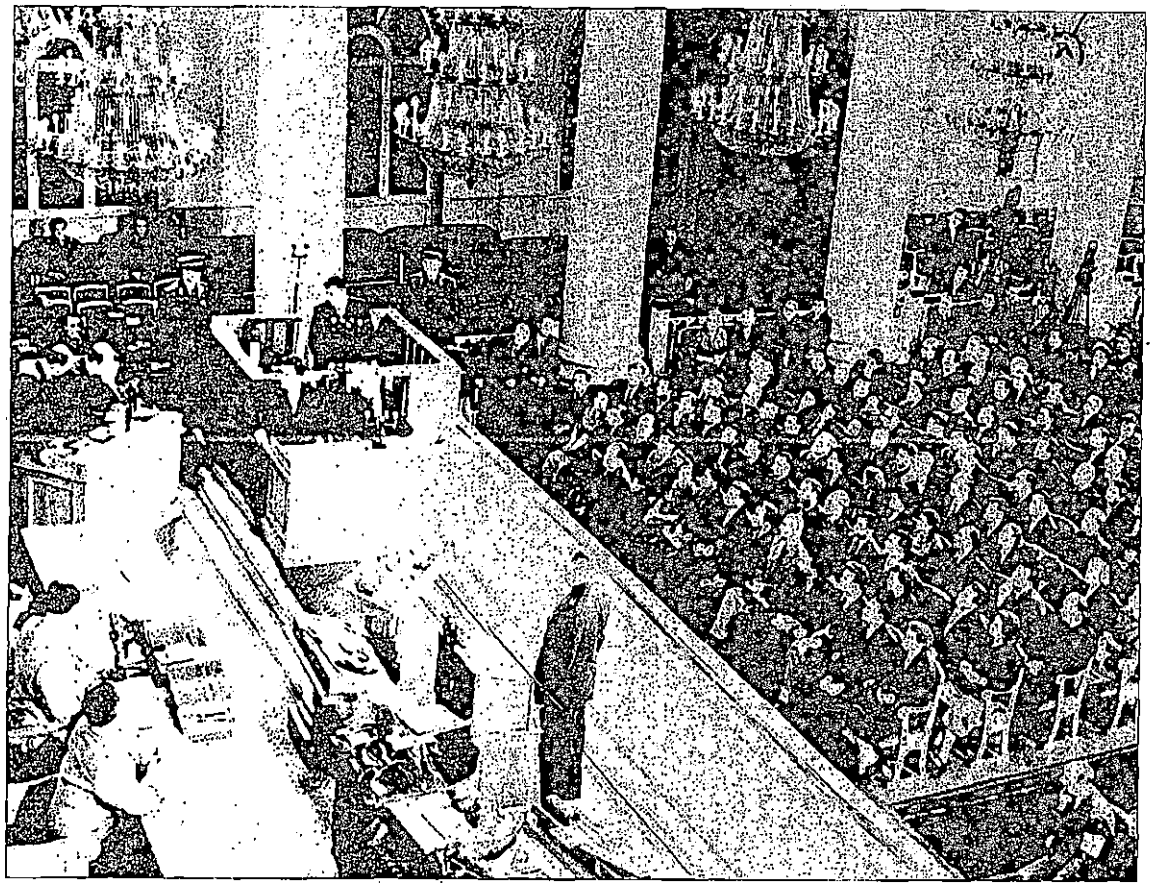
THE FATE OF FRANCIS GARY POWERS

Downed U-2 pilot Francis Gary Powers underwent extensive interrogation at the hands of the Soviets. His instructions from the CIA on what to do in the event of capture were meager, and he had been told that he might as well tell the Soviets whatever they wanted to know because they could get the information from his aircraft anyway. Nevertheless, Powers tried to conceal as much classified information as possible while giving the appearance of cooperating with his captors. To extract the maximum propaganda value from the U-2 Affair, the Soviets prepared an elaborate show trial for Powers, which began on 17 August 1960. Powers continued to conceal as much information as possible, but, on the advice of his Soviet defense counsel, he stated that he was sorry for his actions. The Soviet court sentenced him to 10 years' "deprivation of liberty," with the first three to be spent in prison.[41]

During the next 18 months, confidential negotiations to obtain the release of Powers took place as the United States explored the possibility of trading convicted Soviet master spy Rudolf Abel for Powers. These negotiations were conducted by Abel's court-appointed defense counsel, former OSS lawyer James Donovan, in correspondence with Abel's "wife" (probably his Soviet control) in East Germany. In November 1961, Acting DCI Pearre Cabell wrote to Secretary of State Dean Rusk supporting such a trade, and on 10 February 1962 the actual exchange took place in the middle of the Glienecke Bridge connecting East and West Berlin. As part of the deal, American graduate student Frederick Pryor, who had been jailed in East Germany for espionage, was released at another location.

After Powers returned to the United States, he underwent extensive debriefing, for many questions about his mission remained unanswered. To conduct the debriefing, the Agency immediately reconvened the Damage Assessment Team that had met for two months in the summer of 1960 to estimate what Powers knew about the overflight program and could have told Soviet interrogators. Given Powers' long involvement with the U-2 program, the team had concluded in 1960 that his knowledge was extensive and he had probably revealed most of it to the Soviets. After two weeks of debriefing Powers in February 1962, however, the team found that the damage was much less than had been estimated, and they were quite satisfied

[41] Powers, *Operation Overflight*, pp. 160-192; Beschloss, *Mayday*, pp. 331-335.



Trial of Francis Gary Powers

with Powers' behavior.[42] After reading the debriefing reports, Allen Dulles expressed support of Powers' actions and told Powers, "We are proud of what you have done," but Dulles had already resigned as DCI in November 1961.[43] The new DCI, John A. McCone, demanded a closer look at Powers' actions and set up a Board of Inquiry headed by retired Federal Judge E. Barrett Prettyman. After eight days of hearings and deliberation, the board reported on 27 February that Powers had acted in accordance with his instructions and had "complied with his obligations as an American citizen during this period." The board, therefore, recommended that he receive his back pay.

[42] James J. White, "Francis Gary Powers—The Unmaking of a Hero, 1960-1965," (draft), CIA History Staff, 1974, p. 19 (S).

[43] Powers, *Operation Overflight*, p. 307.

The Prettyman Board's finding was based on a large body of evidence indicating that Powers was telling the truth about the events of 1 May 1960: the testimony of the experts who had debriefed Powers after his return; a thorough investigation of Powers' background with testimony by doctors, psychiatrists, former Air Force colleagues, and his commander at Adana; Powers' own testimony before the board; the results of a polygraph examination that he had volunteered to undergo; and the evidence provided by photographs of the wreckage of his aircraft, which Kelly Johnson had analyzed and found consistent with Powers' story. Nevertheless, DCI McCone remained skeptical. He asked the Air Force to convene its own panel of experts to check Johnson's assessment of the photographs of the U-2. The Air Force quickly complied, and the panel supported Johnson's findings. McCone then seized upon the one piece of evidence that contradicted Powers' testimony—a report by the National Security Agency (NSA) that suggested that Powers may have descended to a lower altitude and turned back in a broad curve toward Sverdlovsk before being downed—and ordered the Prettyman Board to reconvene on 1 March for another look at this evidence. The board remained unconvinced by NSA's thin evidence and stuck to its original findings. A few days later, on 6 March 1962, Powers appeared before the Senate Armed Services Committee, which commended his actions. The Senate Foreign Relations Committee also held brief hearings on the U-2 Affair, with DCI McCone representing the CIA.[44]

Although all of these inquiries found Powers to have acted properly, they did not release many of their favorable findings to the public, which had received a very negative image of Powers' behavior from sensational press reports and statements by public figures who were not aware of (or chose to ignore) the truth about Powers' actions while in captivity. One member of the Senate Foreign Relations Committee, Senator John J. Williams, expressed concern about the impact of this silence on Powers' reputation in a question to DCI McCone on 6 March 1962: "Don't you think he is being left with just a little bit of a cloud hanging over him? If he did everything he is supposed to do, why leave it hanging?"[45] Doubts about Powers did remain in the public mind because he received no public recognition for his efforts to withhold information from the Soviets. He was also

[44] Beschloss, *Mayday*, p. 352-354; Thomas Powers, *Man Who Kept the Secrets*, p. 328; Prettyman Board, DCI records ~~(S)~~.

[45] United States Congress, Senate, Foreign Relations Committee, *Executive Sessions of the Senate Foreign Relations Committee (Historical Series)*, vol. 12, 86th Congress, Second Session, "Report on the U-2 Incident," 6 March 1962, p. 265 (declassified 1982).

snubbed by President Kennedy, who one year earlier had warmly welcomed two Air Force RB-47 fliers released by the Soviet Union. McCone remained hostile to Powers, and in April 1963 he awarded the Intelligence Star to all of the U-2 pilots except Powers. Finally on 25 April 1965, just two days before McCone's resignation became effective, Powers received the Star (which was dated 1963 on the back) from DDCI Marshall S. Carter.[46]

Powers' return from captivity raised the question of what his future employment should be. This issue had already been discussed one year earlier by John N. McMahon, executive officer of the DPD, who noted that he and Col. Leo P. Geary (the Air Force project officer) were concerned about a major dilemma for the CIA and the US Government: "On the one hand we have gone to considerable lengths to prove that the U-2 program was a civilian undertaking and not military aggression; on the other hand there is on file a document that assures Francis Gary Powers that if he so desires he may be *reinstated* into the USAF." On 21 March 1961 McMahon wrote:

> *If we grant him [Powers] the right that is now his, namely reinstatement in the Air Force, then we would be subjecting ourselves to probable adverse propaganda by the USSR. Admitting little appreciation for the finer points of political and psychological warfare, should Francis Gary Powers return to the USAF I suspect that the Soviets would have a "PP" field day illustrating our big lie. The question then, since we cannot permit Powers to return to the USAF, is what do we do with him.*[47]

Despite this negative recommendation, the Air Force agreed on 4 April 1962 to reinstate Powers effective 1 July, a decision that was approved by the Agency, State Department, and White House. Then Powers' divorce proceedings began, and the Air Force, concerned about adverse publicity, postponed reinstatement until the end of the proceedings. In the meantime Powers began working for Lockheed as a U-2 pilot. In March 1963, he met with Colonel Geary to discuss his future plans and decided to stay with Lockheed.[48] Powers remained at Lockheed until U-2 testing ceased in September 1969. Earlier in the year, he had published an account of his experiences on

---

[46] *OSA History*, chap. 14, p. 54 (~~PS~~ Codeword); Beschloss, *Mayday*, p. 397.

[47] John N. McMahon to Chief, Cover Staff, DPD, 21 March 1961, Operation MUDLARK files, OSA records, job 74-B-605, box 6 (~~S~~).

[48] *OSA History*, chap. 14, p. 52 (~~PS~~ Codeword).

the U-2 project under the title *Operation Overflight*. Later he flew a light plane as a traffic reporter for a Los Angeles radio station and then a helicopter for a television station. On 1 August 1977, he and a cameraman from the station died when his helicopter crashed on the way to an assignment.[49]

CHANGES IN OVERFLIGHT PROCEDURES AFTER MAY 1960

One of the most important changes in the overflight program after the loss of Francis Gary Powers' U-2 was the institution of more formal procedures for the approval of U-2 missions. During the first four years of U-2 activity, very few members of the Eisenhower administration had been involved in making decisions concerning the overflight program. The President personally authorized all flights over the Soviet Union and was consulted by Richard Bissell and either the DCI or the DDCI about each such proposed mission. In addition to CIA officials, the President's discussions of individual U-2 missions or of the program as a whole generally included the Secretary of State or his Under Secretary, the Chairman of the Joint Chiefs of Staff, the Secretary of Defense or his deputy, and the President's secretary, Colonel (later General) Goodpaster.

The approval process under President Eisenhower was thus very unstructured. There was no formal approval body charged with reviewing overflight proposals; the President kept this authority in his hands and simply consulted with selected cabinet officials and advisers before reaching a decision. In 1959 the U-2 program had gained a second approval authority when British Prime Minister Harold Macmillan became the approval authority for missions conducted by the RAF pilots in Detachment B.

The loss of Powers' U-2 in May 1960 led to major changes in the approval process. For all practical purposes, Prime Minister Macmillan ceased to be a source of approval because the RAF pilots who remained in the U-2 program did not conduct any more operational missions (although the use of British pilots was considered on several occasions). In the United States the approval process

[49] Beschloss, *Mayday*, pp. 396-401. Beschloss claims that Powers was fired by Lockheed for criticizing the Agency in his memoirs (which he had shown to the Agency in draft form), but Kelly Johnson's "U-2R Log" records on 25 September 1969: "We have no flight test activity at all. I must let Gary Powers go. Have protected him for about seven years, but he doesn't have an ATR (Air Transport Rating), so we have no other job for him—not even flying the Beechcraft."

became more formal as the National Security Council became involved. Henceforth, proposed missions had to be submitted to the National Security Council (NSC) Special Group for approval. In the early 1960s, the Special Group consisted of the DCI, the Deputy Secretary of Defense, the Under Secretary of State, and the Military Adviser to the President. After the Military Adviser, Gen. Maxwell Taylor, became Chairman of the Joint Chiefs of Staff in 1962, his place on the Special Group was taken by McGeorge Bundy, the President's Special Assistant for National Security Affairs.[50]

Before requesting permission from the Special Group for a U-2 mission over denied territory, the CIA prepared a detailed submission giving justification for the proposed mission and maps showing the targets to be photographed, flight times, and emergency landing sites. Such submissions came to be known as "black books" because they were placed in black, looseleaf binders. The decision of the Special Group was generally final, although on occasion controversial issues were presented to the President for his decision.

This approval process did not come into play immediately after May 1960 because there was a long pause in U-2 operations as the detachments returned from overseas. It was not until late October 1960 that the next U-2 operation occurred, this time over Cuba. By this time the full approval procedure had been established, and the Special Group approved the mission (see chapter 5).

The approval process was not the only part of the U-2 program that changed after May 1960. The process for establishing requirements for overhead reconnaissance missions also became more formal. In August 1960 the US Intelligence Board took over the Ad Hoc Requirements Committee and merged it with the Satellite Intelligence Requirements Committee to form the Committee on Overhead Reconnaissance. DCI Directive 2/7 tasked COMOR with the "coordinated development of foreign intelligence requirements for overhead-reconnaissance projects over denied areas." The DCID defined "overhead reconnaissance" to include "all reconnaissance for foreign-intelligence purposes by satellite, or by any vehicle over

[50] The Special Group, which had been created by NSC Intelligence Document 5412/2 in 1955 to oversee covert activities, was originally known as the 5412 Committee. Later the Special Group became known as the 303 Committee and then the 40 Committee. United States Congress, Senate, Select Committee to Study Governmental Operations with Respect to Intelligence Activities, *Foreign and Military Intelligence*, book 1, (Washington, DC: US Government Printing Office, 1976), pp. 48-53.

denied areas, whether by photographic, ELINT, COMINT, infrared, RADINT, or other means." The only exception to COMOR's area of responsibility was "reconnaissance and aerial surveillance in direct support of actively combatant forces."[51]

By this time the Air Force had developed a large overhead reconnaissance program of its own, including a fleet of U-2s, and, occasionally, there were conflicts between the areas of responsibility of COMOR and the military services for collection requirements. The Air Force had already won a major victory in 1958, when it claimed that the White House had given responsibility for peripheral reconnaissance of the Soviet Union to the military. DCI Dulles, who was always reluctant to become involved in matters that seemed to lie in the military's area of responsibility, did not resist this claim, and the Ad Hoc Requirements Committee stopped preparing requirements for peripheral flights. This ended a major requirements committee study, which sought to estimate what could be gained from U-2 oblique photography along the entire border of the Soviet Union.[52] The last CIA U-2 mission along the Soviet Union's coasts occurred on 22 June 1958; thereafter, the only peripheral missions conducted by the CIA were those along the Soviet Union's southern border with Iran and Afghanistan from bases in Pakistan and Turkey under covert arrangements with the host governments.

Until the spring of 1961, there was virtually no coordination of military reconnaissance activities, even within the individual services. Each commander of a Theater or a Unified and Specified Command conducted his own independent reconnaissance activities. To meet the growing need for overall coordination of these activities at the national level, the Joint Chiefs of Staff (JCS) established the Joint Reconnaissance Center (JRC) under the J-3 (Operations) of the Joint Staff. The JRC immediately began to coordinate and obtain approval for approximately 500 missions per month, assigning each a risk factor of Critical, Sensitive, Unique, or Routine. The JRC then prepared a monthly Activities Book giving details of the proposed missions and briefed the Joint Chiefs of Staff on the more risky missions. The CIA received a copy of the Activities Book.

---

[51] DCID 2/7, effective 9 August 1960 (~~S~~).

[52] Memorandum for DCI McCone from James Q. Reber, Chairman, COMOR, "Proposed Procedures for Approval of Critical Reconnaissance," 21 March 1962, COMIREX records (~~TS~~ Codeword).

Most military reconnaissance missions were approved or disapproved at the JCS level, but the most sensitive missions were submitted through the Secretary of Defense to the Special Group for approval. In addition to this Department of Defense approval path, the military services could also submit requirements through the DCI using their representatives on COMOR. As a result, the military services had two channels for submitting reconnaissance missions to the Special Group. The Agency had only one—COMOR.[53]

The main conflicts between the requirements committee and the military services arose over missions in the Far East. In the early 1960s, North Vietnam had not been designated a denied area by the US Intelligence Board (USIB), so the military services could plan missions there without consulting COMOR. Such missions, however, came very close to China, which was a denied area and, therefore, came under COMOR's area of responsibility. Once the war in Southeast Asia escalated in 1964, the military services received responsibility for the entire area (see chapter 5).

To reduce the number of disputes between the competing CIA and Air Force reconnaissance programs and to manage the growing satellite program, the two agencies worked out an agreement to provide overall coordination for reconnaissance activities at the national level. The first such interagency agreement came in the fall of 1961, and it was followed by three additional agreements during the next four years.[54]

Interest in coordinating the reconnaissance efforts of the military services and the CIA also affected the field of photographic interpretation. In the wake of the loss of Francis Gary Powers' U-2 on 1 May 1960, the President's Board of Consultants on Foreign Intelligence Activities (PFIAB) had urged the establishment of an interagency group to study ways to improve the entire US intelligence community. Formed on 6 May 1960, the Joint Study Group on Foreign Intelligence Activities met for the next seven months under the leadership of Lyman Kirkpatrick, CIA Inspector General. One of the study group's key recommendations in the report it issued in December 1960 was the creation of a national photointerpretation

[53] Ibid (~~CVS~~ Codeword).

[54] Problems of classification prevent a more detailed discussion of this aspect of the reconnaissance program, which will be covered in a future history of satellite reconnaissance at a higher level of classification.

center that would bring together photointerpreters from the Agency and the military services. The report further recommended that the CIA be placed in charge of the new center. Ignoring Air Force claims that it should head such a center, President Eisenhower approved the report's recommendation, and, on 18 January 1961, National Security Council Intelligence Directive (NSCID) No. 8 established the National Photographic Interpretation Center (NPIC). Henceforth, the director of NPIC would be designated by the DCI and approved by the Secretary of Defense, and the deputy director would come from one of the military services. The first director of NPIC was Arthur S. Lundahl, head of the CIA's Photo-Intelligence Division.[55]

One additional major change in the U-2 program in the years immediately following the May Day incident—although not directly related to the loss of Powers' U-2—was the departure of Richard Bissell from the CIA and the subsequent reorganization of the Agency's reconnaissance and scientific activities. The roots of Bissell's downfall went back to 1 January 1959, when he became Deputy Director for Plans and decided to place all Agency air assets in the DDP in order to maintain control of his overhead reconnaissance projects (the U-2 and its two proposed successors, the OXCART aircraft and the reconnaissance satellite). The previously independent Development Projects Staff became the Development Projects Division (DPD) of the DDP and now controlled all Agency air operations, including air support for covert operations. As a result, U-2s were occasionally employed for gathering intelligence to support DDP operations in addition to their primary mission of gathering strategic and tactical intelligence.

Although the reorganization made sense in terms of increasing the efficiency of Agency air operations, the use of the U-2 to support covert action disturbed Bissell's backers among the scientists advising Presidents Eisenhower and Kennedy, especially James Killian and Edwin Land. They were concerned that Bissell was becoming too involved in covert action and was not able to devote sufficient time to the overhead reconnaissance program. Then came the disastrous Bay of Pigs invasion in April 1961, which discredited Bissell with the Kennedy administration in general and the two scientists in particular. Later that year, Bissell lost another important source of support when Allen Dulles resigned as DCI in November 1961. During his final

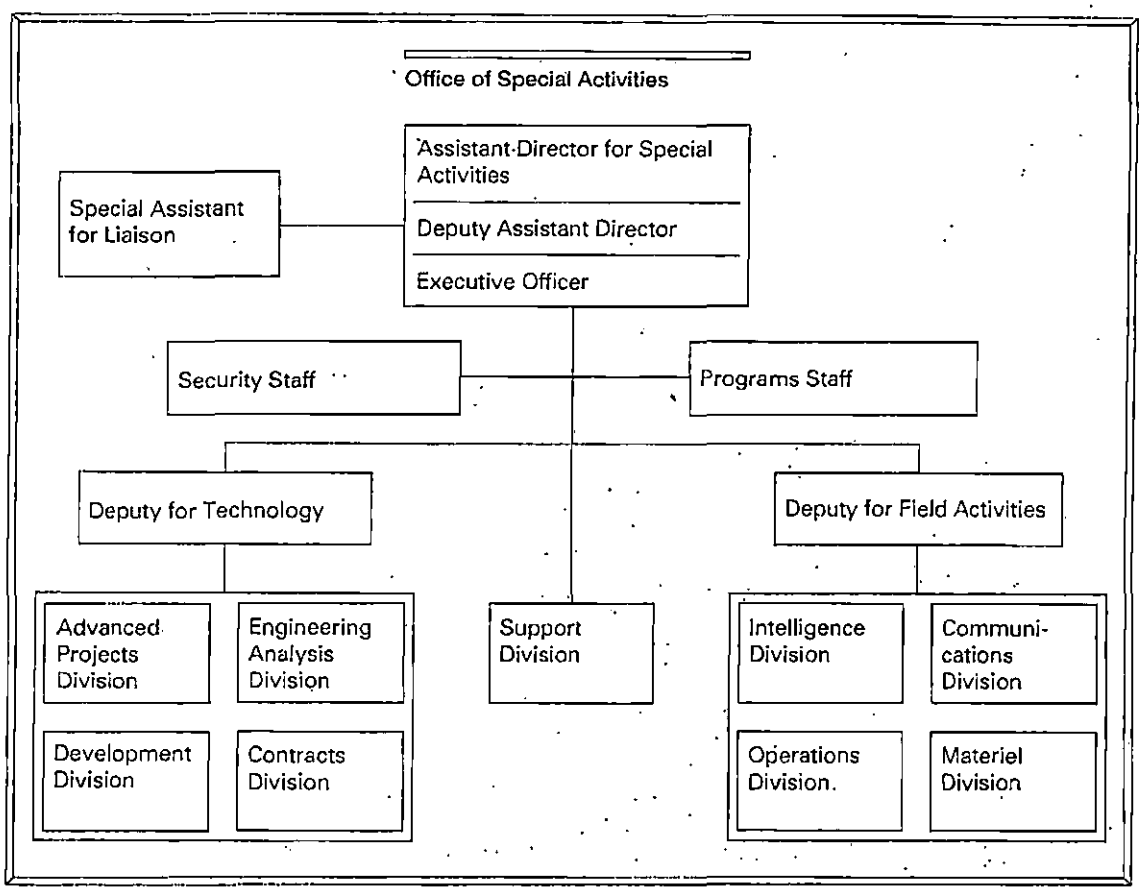
[55] Lundahl and Brugioni interview (~~TK~~ Codeword).

months as the Deputy Director for Plans, Bissell found himself involved in a major struggle with Killian and Land, who were serving on President Kennedy's Foreign Intelligence Advisory Board (successor to the Eisenhower administration's President's Board of Consultants on Foreign Intelligence Activities). These two influential Presidential advisers strongly advocated removing the Agency's overhead reconnaissance programs from the DDP and placing them in a new, science-oriented directorate, but Bissell resisted this proposal. With his position in the Agency becoming increasingly untenable, Bissell resigned on 17 February 1962, after turning down an offer from the new DCI, John A. McCone, to become the CIA's first Deputy Director for Research.[56]

Two days after Bissell's departure, the new Directorate came into existence, and it absorbed all of the Development Projects Division's special reconnaissance projects. Only conventional air support for the Clandestine Services remained with the DDP in the new Special Operations Division. The U-2 program was no longer connected with covert operations.

The first half of 1962 was a confusing period for the Development Projects Division. After losing the individual who had created and supervised it for seven years, the DPD also lost its feeling of autonomy when it was transferred from its own building to the new CIA Headquarters at Langley. Soon afterward, Col. Stanley W. Beerli, who had headed the DPD since 1960, returned to the Air Force. Then on 30 July 1962, the overhead reconnaissance projects underwent a major reorganization with the formation of the new Office of Special Activities (OSA) to replace the DPD. The original organization of OSA with 10 division or staff heads reporting directly to the director of the office (at that time known as the Assistant Director for Special Activities) proved too cumbersome, and, on 30 September 1962, a reorganization divided most of these offices between two major subordinates, the Deputy for Technology and the Deputy for Field Activities (see chart, page 193). The Office of Special Activities (OSA) continued to control reconnaissance activities and related research and development after the Directorate of Research was enlarged and renamed the Deputy Directorate for Science and Technology (DDS&T) on 5 August 1963 (along with the other

[56] Killian interview (S); Land interview (TS Codeword); Richard M. Bissell to John A. McCone, 7 February 1962, DCI records, job 80-B-1676R, box 18, folder 10 (S).



Directorates, DDS&T dropped the "Deputy" from its title in 1965 and became known as the Directorate of Science and Technology). In 1965 the head of OSA received a new title, Director of Special Activities. The Office of Special Activities remained in control of the CIA's overhead reconnaissance activities until 1974, when the Agency ended its involvement with manned reconnaissance aircraft.[57]

[57] *OSA Chronology*, pp. 34-35 (~~PS~~ Codeword).

U-2 Operations After May 1960

The loss of Francis Gary Powers' U-2 over the Soviet Union on 1 May 1960 marked the end of the aircraft's use over the Soviet Bloc. Soon after the May Day incident, President Eisenhower ordered an end to overflights. Similarly, his successor, John F. Kennedy, told a 25 January 1961 press conference, "I have ordered that the flights not be resumed, which is a continuation of the order given by President Eisenhower in May of last year." This was not a binding pledge, as John A. McCone (who became DCI in November 1961) pointed out to President Kennedy's successor, Lyndon B. Johnson, on 15 January 1964 in response to the new President's request for information on U-2 overflight policies:

> *Contrary to popular assumption, President Kennedy did not make any pledge or give an assurance, at least publicly, that there would be no further overflights. He limited his response to a statement that he had ordered that the flights not be resumed. An order, obviously, is valid only until countermanded.*[1]

Technically, McCone was correct, but no President was likely to order a resumption of overflights of the Soviet Union without very good reason, and such a situation never developed, in part because satellite photography gradually began to fill the gap left by the end of U-2 coverage.

Although there were several proposals to resume overflights of the Soviet Union in the years that followed, none reached the mission planning stage. The Kennedy administration came closest to resuming

[1] Memorandum for President Johnson from DCI McCone, "Response to Query Concerning U-2 Overflight Policy," 15 January 1964, DCI records, job 80-B-1676R, box 17, folder 14 (~~EPS~~ Codeword).

overflights of the Soviet Union during the Berlin Crisis in the summer and fall of 1961. On 14 September 1961, Kelly Johnson noted in his project log:

> *Have had request from Mr. Bissell to propose ways and means for increasing safety of the U-2 on probable overflights.... It seems that President Kennedy, who publicly stated that no U-2's would ever be over Russia while he was president, has requested additional flights. Some poetic justice in this.*[2]

One week later Colonel Geary called to order Lockheed to upgrade six older U-2s into U-2Cs with the more powerful engines on a priority basis, even if it meant taking people off the work on the successor aircraft in order to speed up the conversions.

Shortly thereafter, the resumption of overflights became a major topic of discussion within the intelligence community. On 25 September 1961, the Committee on Overhead Reconnaissance prepared a detailed "Justification for U-2 Photography over the USSR," which argued in favor of U-2 missions over selected, high-priority targets such as ICBM complexes. The COMOR paper stated that satellite photography did not provide sufficient detail to answer many critical questions about the Soviet ICBM program. To back up this contention, the report placed U-2 and satellite photography of the same Soviet targets side by side, clearly demonstrating the far superior resolution of the U-2's cameras. Not all members of COMOR supported the resumption of overflights, however. When COMOR formally recommended this course of action to the USIB on 1 October 1961, the State Department and CIA members dissented, having found "insufficient justification for resuming U-2 overflights of the USSR at this time."[3]

---

[2] Johnson, "Log for Project X," 14 September 1961. In preparation for the possible resumption of overflights, Kelly Johnson began thinking about what to do in a worst case scenario like that of 1 May 1960. He noted in the project log on 21 September 1961: *One of the greatest technical problems and, of course, a great moral one, is how we insure destroying the aircraft and the pilot should the mission fail. I have proposed a time-altitude fusing setup for multitude bombs, that looks like it should do the trick. Beerli [Col. Stanley Beerli, USAF, Director of the Office of Special Activities] doesn't want anything to do with this, but we will go ahead and develop it in case someone decides it is necessary.*

[3] Memorandum for USIB from COMOR, "Justification for U-2 Photography over the USSR," 25 September 1961, IC Staff, COMIREX records, job 33-T-123A, box 10, "COMOR (General)" (TS Codeword); Memorandum for USIB from COMOR, "Requirements for Resumption of U-2 Overflights of the USSR," 1 October 1961, IC Staff, COMIREX records, job 33-B-119A, box 1 (TS Codeword).

Nothing came of the proposal to resume overflights in the fall of 1961, as both the USIB and the Special Group came out against it, but, as long as U-2 photography remained clearly superior to satellite photography, the thought of obtaining U-2 coverage of the Soviet Union remained tempting. In February 1962, the USIB seriously considered a COMOR proposal to send a U-2 over Kamchatka to photograph Soviet antiballistic-missile facilities but finally decided to wait for the results of an Air Force peripheral mission. The board later accepted DCI McCone's recommendation to seek satellite rather than U-2 coverage of the area.[4]

With both the CIA and the State Department strongly opposed to sending the highly vulnerable U-2 over the Soviet Union, prospects for resuming flights remained slight unless the international situation worsened to such a degree that overflights would be worth the risks involved. Since this never happened, Francis Gary Powers' flight on 1 May 1960 proved to be the last CIA overflight of the Soviet Bloc. Yet, the U-2 remained useful, for it could operate successfully in other areas with less developed radar and air defense systems. After May 1960, the main focus of U-2 activity shifted to two new areas: Latin America, where U-2s would play an extremely important role during the early 1960s, and the Far East, where CIA U-2s were active from 1958 until 1974, when the Agency's involvement in manned reconnaissance finally ended.

U-2 OPERATIONS IN LATIN AMERICA

U-2 Support to the Bay of Pigs Invasion

During late summer 1960, the Directorate of Plans was planning a counterrevolutionary invasion of Cuba for the following year. To support this effort, the Agency asked the National Security Council's

[4] Memorandum for the Special Group from COMOR, "Illustrations of Policy Restraints on the Collection of Information through Overflight of Denied Areas during 1962," 14 December 1962, IC Staff, COMIREX records, job 33-B-119A, box 1 (~~TS~~ Codeword); James S. Lay, "The United States Intelligence Board, 1958-1965," (draft) CIA History Staff MS-2, 1974, p. 385 (~~TS~~ Codeword). One year later Saryshagan was the topic of US Intelligence Board deliberations. In October 1963 the board asked COMOR to prepare recommendations on the need for an electronic intelligence-gathering mission against the Soviet ABM installations at Saryshagan. The proposed mission would not, however, violate Soviet airspace; instead, the U-2 would fly over the portion of the People's Republic of China closest to Saryshagan. Lay, "USIB History," pp. 393-94 (~~TS~~ Codeword).

Special Group to approve U-2 overflights of Cuba. Known as Operation KICK OFF, these flights were designed to obtain intelligence on Cuban air and ground order of battle and to provide geographic data for choosing an invasion site.

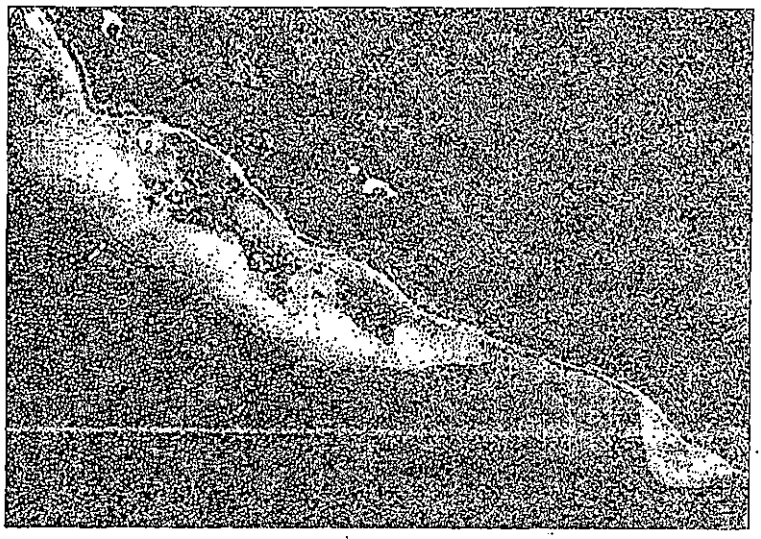
To allay fears that mechanical problems could lead to the loss of a U-2 over Cuba, the submission to the Special Group for overflights emphasized that, if a U-2 had a flameout anywhere over Cuba, it could still glide back and make a safe landing in Florida. The Special Group approved Operation KICK OFF but stipulated that only two overflights could be made. Detachment G staged the Cuban missions from Laughlin AFB near Del Rio, Texas, a base used by SAC U-2 aircraft. Agency photointerpreters went to Del Rio to read out the photography after these missions. The two flights, on 26 and 27 October 1960, were very long missions, covering 3,500 miles and lasting over nine hours. Because of cloud cover over Cuba, the results of both missions were poor. The Agency, therefore, asked the Special Group to approve additional missions. After receiving authorization, Detachment G conducted three missions (Operation GREEN EYES) on 27 November and 5 and 11 December 1960 with good results.

Overflights of Cuba continued under the new administration of President Kennedy. Under the codename Operation LONG GREEN, two overflights on 19 and 21 March 1961 photographed Cuba extensively to aid the final preparations for the invasion. Two weeks later Detachment G again deployed from Edwards AFB, California, to Laughlin AFB, Texas. Beginning on 6 April, Detachment G U-2s made 15 flights over Cuba to provide photographic coverage of the ill-fated Bay of Pigs invasion and its aftermath. These flights were known as Operation FLIP TOP.[3]

Aerial Refueling Capability for the U-2

Long missions conducted over Cuba in late 1960 and over Southeast Asia in early 1961 pointed out the need to increase the range of the U-2. In May 1961, Lockheed began modifying Agency U-2s so that they could be refueled in flight to extend their operating range. The six Agency aircraft that were modified to achieve this capability received the designation U-2F. All Agency U-2 pilots then underwent training in the techniques of in-flight refueling.

[3] *OSA History*, chap. 16, pp. 13-15 (~~CRS~~ Codeword).



Aftermath of the Bay of Pigs invasion, 20 April 1961

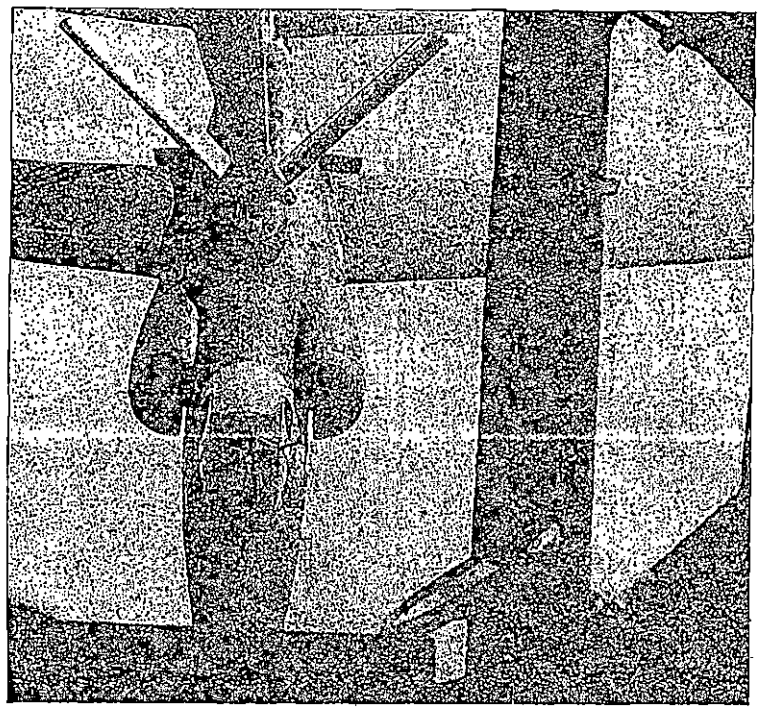
Refueling a U-2 in flight was a very delicate task. When fully loaded with fuel, KC-135 tankers found it difficult to reduce airspeed to 200 knots, the safest speed for refueling a U-2. As for the U-2s, they were in a very vulnerable position when approaching a tanker at 200 knots because their frail wings could not stand much stress. As a result, U-2 pilots had to approach the KC-135 tankers very carefully in order to avoid the vortexes from the wingtips of the tanker and the turbulence caused by the four large jet engines. During the first few years of refueling operations, two U-2s crashed after their wings broke off as they crossed into the turbulent area behind the tankers; one of the pilots was killed.[6]

The in-flight refueling capability was a useful modification to the U-2, but it could not dramatically extend mission length. The main limiting factor remained pilot fatigue, which prevented missions from lasting longer than approximately 10 hours.

U-2 Coverage During the Cuban Missile Crisis

Cuba remained a high-priority target even after the Bay of Pigs invasion failed in April 1961. Soon afterward, Detachment G U-2s began flying monthly missions over Cuba in a program known as Project

[6] Ibid., p. 11-12 (~~US~~ Codeword).



In-flight refueling of a U-2

NIMBUS. Most of the flights were staged from Laughlin AFB, Texas, but three were flown from Edwards AFB, California, using in-flight refueling to extend the range of the aircraft. By the spring of 1962, having received reports of increased Soviet activity in Cuba, the CIA requested permission for additional photographic coverage of the island. The Special Group authorized increasing the number of Cuban overflights to at least two per month, beginning in May 1962. At the same time, the National Photographic Interpretation Center began publishing a *Photographic Evaluation of Information on Cuba* series.[7]

By early August 1962, CIA analysts had noted a substantial increase in Soviet arms deliveries to Cuba during the preceding weeks. The first U-2 overflight in August, mission 3086 on the 5th, flew too soon to detect the Soviet construction program just getting under way at various sites in Cuba. A second mission (3088) was originally set for 8 August, but bad weather forced repeated postponements until 29 August. This mission's photography provided the first hard evidence

[7] Ibid., pp. 19-20 (~~TS~~ Codeword).

of the nature of the Soviet buildup in Cuba. Two days after the mission, the CIA reported in the *President's Intelligence Checklist* that there were at least eight surface-to-air missile (SA-2) sites in the western half of Cuba.[8] (The map on page 202 shows the routes taken by the two August overflights.)

DCI John A. McCone

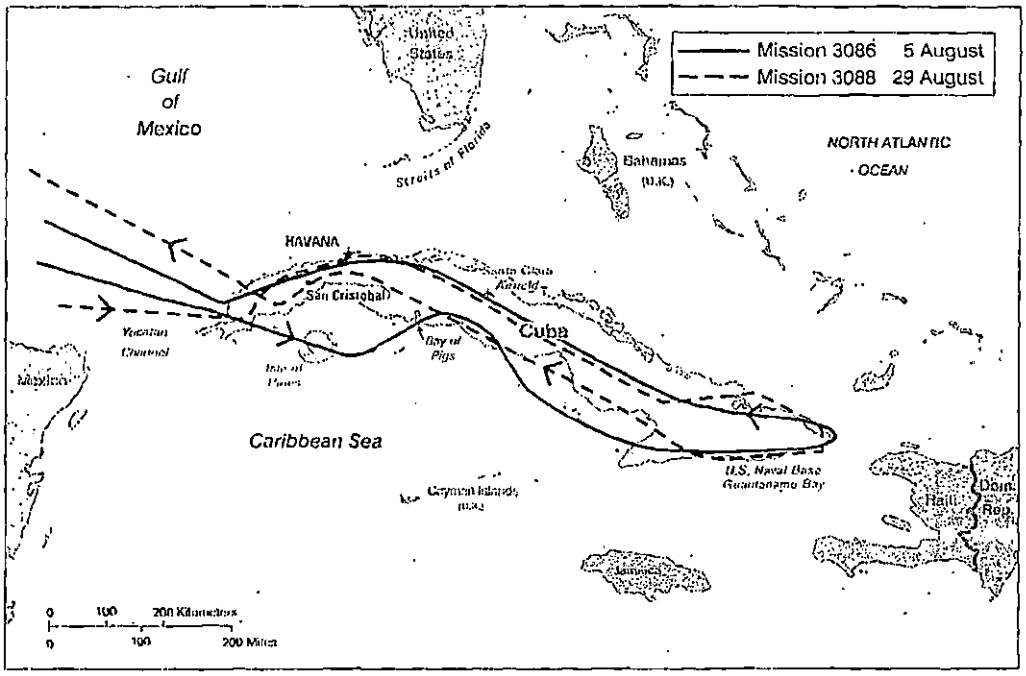
On 5 September the next U-2 overflight (mission 3089) provided more evidence of the Soviet buildup. The mission's photography showed three more SAM sites and also revealed a MiG-21, one of the newest Soviet fighter aircraft, at the Santa Clara airfield.

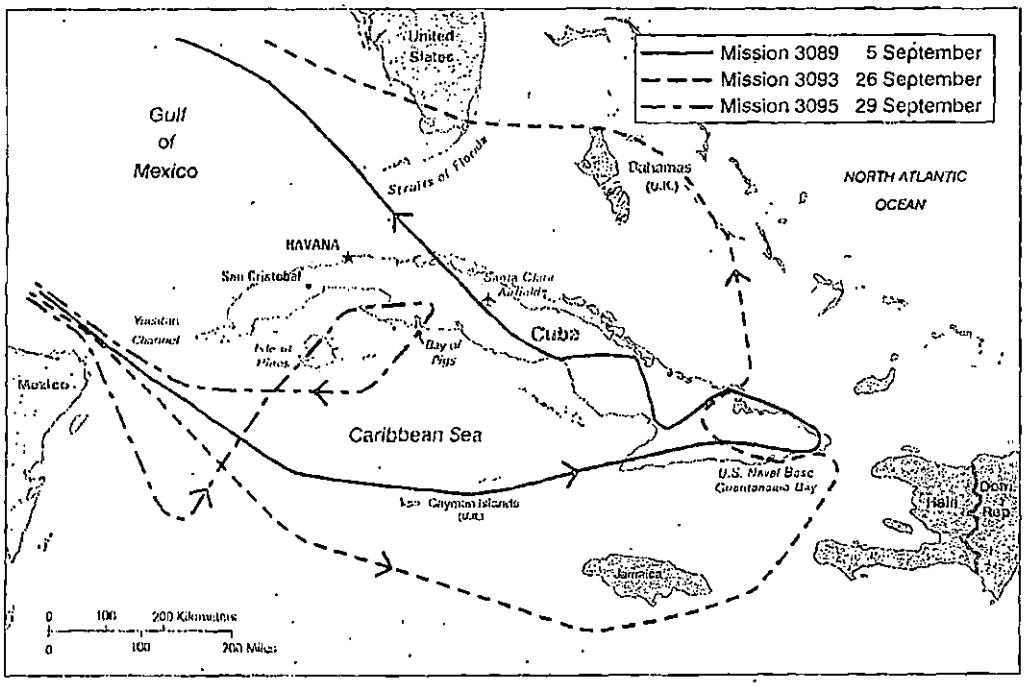
The discovery of SAMs in Cuba had a twofold effect on the US reconnaissance effort over Cuba. First, it added substance to DCI McCone's fears that Cuba might become a base for Soviet medium-range ballistic missiles (he argued that SAM sites would only be set up to protect high-priority facilities such as missile bases). At this time, however, McCone's suspicions were not shared by other officials in the Agency or the administration. The second and most significant effect of the discovery of SAMs in Cuba was to make the administration far more cautious in its use of U-2s for reconnaissance of the island. As the loss of Francis Gary Powers' U-2 in May 1960 had demonstrated, the U-2 was very vulnerable to the SA-2 missile.

Within the administration, concern mounted about the U-2's vulnerability to SAMs in Cuba and the possibility that a loss could cause a major diplomatic crisis. Such fears increased as the result of two incidents in other parts of the world. On 30 August 1962, a SAC U-2 on a peripheral reconnaissance mission overflew Sakhalin Island in the Far East, prompting a Soviet protest on 4 September. The United States apologized for the intrusion. Then on 8 September, a U-2 with a Nationalist Chinese pilot was shot down over the People's Republic of China (this CIA reconnaissance program is discussed later in this chapter in the section on Asian operations). Increasing concern about U-2 vulnerability led to an impromptu meeting on 10 September 1962 of Secretary of State Dean Rusk, National Security Adviser McGeorge Bundy, and DDCI Marshall S. Carter (in place of the DCI, who was on his honeymoon in France). The Secretary of State objected to the CIA's plans for two extended overflights covering the remaining areas of Cuba not covered by the last two missions. Rusk wanted peripheral flights over international waters kept separate from

[8] Richard Lehman, "CIA Handling of the Soviet Buildup in Cuba, 1 July–16 October 1962," 14 November 1962 (Hereafter cited as Lehman Report), DCI records, job 80-B-1676R, box 17, folder 18 (~~OXCART~~ Codeword).

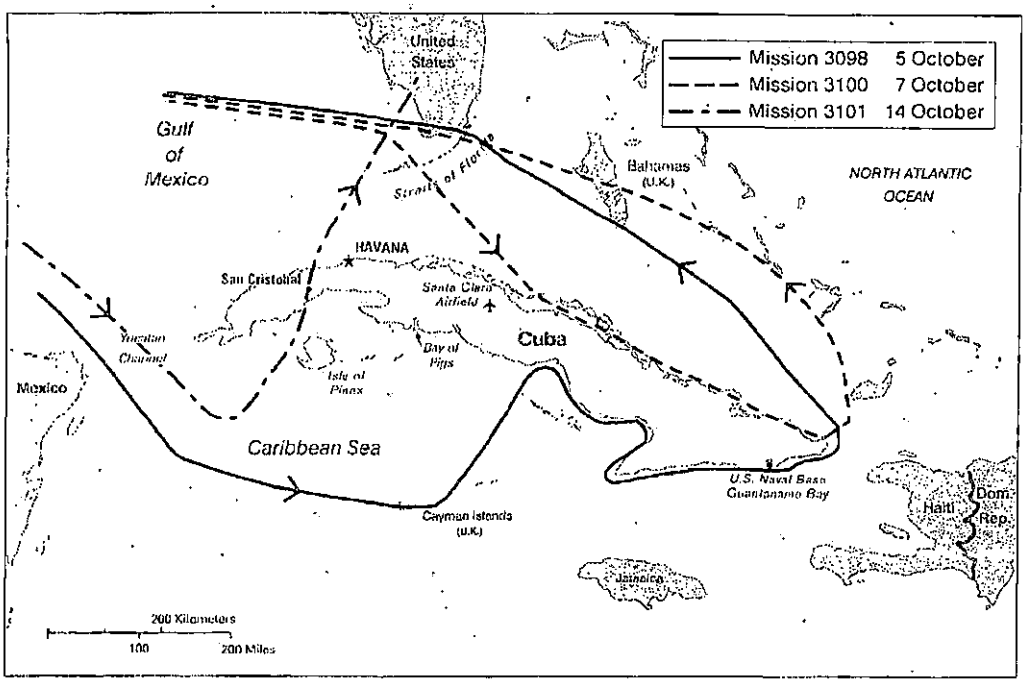
U-2 Overflights of Cuba, August - October 1962

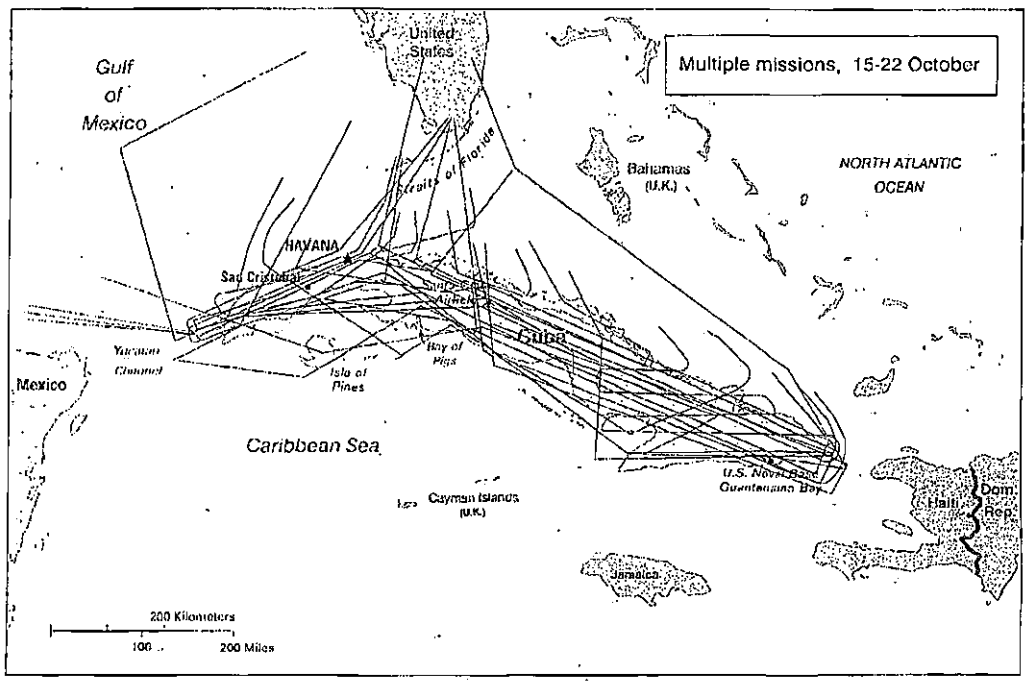




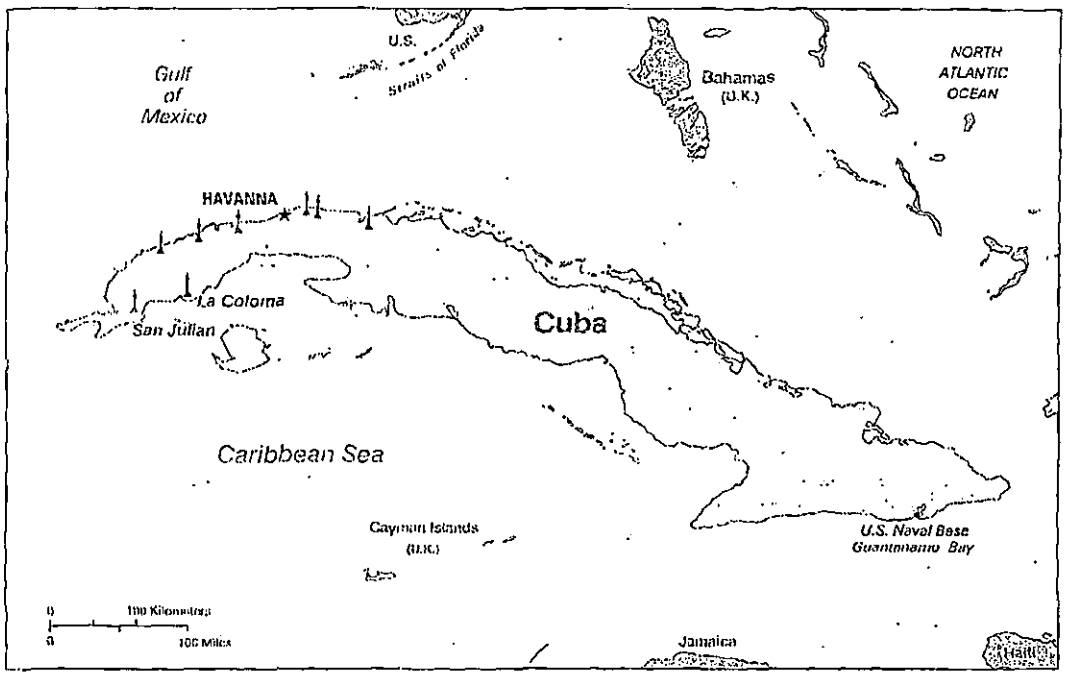
~~Secret~~ NOFORN

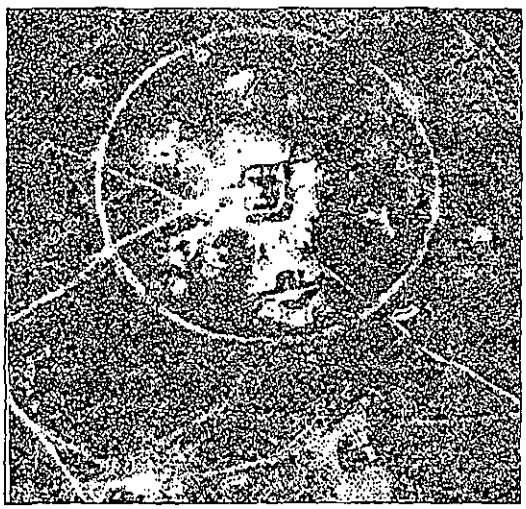
724752 (R00426) 4-92


Mission 3098 5 October
Mission 3100 7 October
Mission 3101 14 October
United States
Gulf of Mexico
Straits of Florida
Bahamas (U.K.)
NORTH ATLANTIC OCEAN
HAVANA
San Cristobal
Santa Clara Airfield
Cuba
Yucatan Channel
Bay of Pigs
Isle of Pines
Mexico
Caribbean Sea
U.S. Naval Base Guantanamo Bay
Haiti
Dom. Rep.
Cayman Islands (U.K.)
Jamaica
200 Kilometers
100
200 Miles

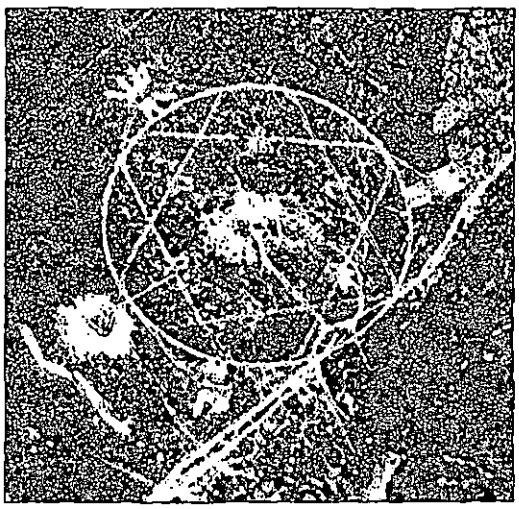

Multiple missions, 15-22 October
United States
Gulf of Mexico
Straits of Florida
Bahamas (U.K.)
NORTH ATLANTIC OCEAN
HAVANA
San Cristobal
Santa Clara Airfield
Cuba
Yucatan Channel
Bay of Pigs
Isle of Pines
Mexico
Caribbean Sea
U.S. Naval Base Guantanamo Bay
Haiti
Dom. Rep.
Cayman Islands (U.K.)
Jamaica
200 Kilometers
100
200 Miles

SAM Sites in Cuba, August 1962





San Julian Airfield



La Coloma

~~Secret~~ NOFORN

724754 (R00420) 4-62

overflights of Cuban territory. He argued that the loss of an aircraft on a mission that combined both types of flights would make it difficult for the United States to stand on its rights to fly over international waters. Bundy and Carter therefore agreed to split the proposed reconnaissance program into four missions: two overflights and two peripheral flights, all planned for maximum safety. The overflights were thus designed to be quick "in-and-out" operations across the narrow width of the island instead of flights along the entire length of Cuba, as had been the case previously. (As the map on page 202 illustrates, the 5 September mission was the last one to fly along the length of the island.) As an additional precaution, flightpaths would be laid out to avoid known SAM sites. Although these changes greatly reduced the danger to the U-2, they slowed the gathering of information on the Soviet buildup by reducing each mission's coverage.[9]

To ensure that the photographs taken by these missions were of the highest quality, the CIA decided to conduct flights only when the weather along the flight routes was less than 25 percent overcast. Weather proved to be a major problem during the month of September. Unfavorable forecasts (along with a brief standdown of U-2 overflights after the loss of the Nationalist Chinese U-2) prevented the launching of any missions from 6 through 16 September. Moreover, when mission 3091 finally flew on 17 September, the favorable weather forecast proved inaccurate and heavy clouds prevented the mission from obtaining usable photography. Bad weather continued to rule out missions until 26 September, when mission 3093 covered eastern Cuba and found three additional SAM sites. Three days later mission 3095 flew over the Isle of Pines and Bay of Pigs area, finding one more SAM site and a coastal-defense cruise missile site.[10]

The cautious series of U-2 flights in September had turned up many more SAM sites but no concrete evidence of the presence of surface-to-surface missiles. Growing impatient with the restrictions

---

[9] Lehman Report, pp. 12-13 (TS Codeword).

[10] DCI John A. McCone, Memorandum for the Record, "U-2 Overflights of Cuba, 29 August through 14 October 1962," 27 February 1963, DCI records, job 80-B-1676R, box 17, folder 18 (S). Although this DCI memo states that "the delay in completing the photographic coverage was due solely to the unfavorable weather predicted during this period," a more contemporary COMOR memo reported a standdown of U-2 overflights until 16 September as a result of the loss of mission No. GRC-127 over China on 8 September. Memorandum for DDCI Carter from James Q. Reber, Chairman, COMOR, "Historical Analysis of U-2 Overflights of Cuba," 24 October 1962, IC Staff, COMIREX records, job 33-B-122A, box 1, "Cuba Requirements, 1961-63" (TS Codeword).

that had been placed on U-2 overflights of Cuba, DCI McCone told the Special Group on 4 October 1962 that their policy of avoiding SAM sites had restricted the Agency to using the U-2 only in Cuba's southeastern quadrant. He questioned "whether this was a reasonable restriction at this time, particularly since the SAM's were almost certainly not operational."[11] The Special Group then requested the preparation of an overall program for reconnaissance of Cuba in time for its next meeting on 9 October.

In the meantime, CIA U-2s continued the reconnaissance program that the Special Group had approved in September. In early October two peripheral missions—3098 along the southeastern coast on 5 October and 3100 along the northern coast on 7 October (see map on page 203)—discovered an additional five SAM sites. This brought the total to 19, but there was still no evidence of surface-to-surface missiles.

Evidence was mounting that the portion of Cuba that the September and early October missions had avoided was the most likely location for Soviet medium-range ballistic missiles (MRBMs). On 6 October 1962, the Committee on Overhead Reconnaissance recommended frequent and regular coverage of Cuba, pointing in particular to the need for renewed coverage of western Cuba:

> *The absence of coverage of the western end since August 29, coupled with the rate of construction we have observed, means that there may well be many more sites now being built of which we are unaware. Ground observers have in several recent instances reported sightings of what they believe to be the SS-4 (SHYSTER) MRBM in Cuba. These reports must be confirmed or denied by photo coverage.*[12] *Attached to this memorandum was a list of targets, with the area around San Cristobal at the top.*

On 9 October the Special Group met to discuss COMOR's recommendations, the most important of which was a U-2 flight over the "suspect MRBM site as soon as weather permits." This mission was also designed to pass over one of the SA-2 sites that was thought to be most nearly operational in order to determine the status of SA-2

[11] Minutes of the Special Group meeting, 4 October 1962, in Memorandum for DCI McCone from J. S. Earman, Inspector General, "Handling of Raw Intelligence Information During the Cuban Arms Buildup," 20 November 1962, DCI records, job 80-B-1676R, box 17 (~~TS~~ Codeword).

[12] Lehman Report, p. 30 (~~TS~~ Codeword).

defenses of Cuba. If this overflight did not provoke an SA-2 reaction, the study recommended "maximum coverage of the western end of the island by multiple U-2s simultaneously."[13] Because the danger posed by the SA-2 sites was one of the major topics at the Special Group meeting, DCI McCone brought along Col. Jack C. Ledford (USAF), head of the Office of Special Activities, who presented a vulnerability analysis that estimated the odds of losing a U-2 over Cuba at 1 in 6. The Special Group approved the recommended flight over San Cristobal.

As the Special Group meeting was breaking up, Deputy Secretary of Defense Roswell Gilpatric and the Air Force representative questioned the adequacy of the Agency's cover story, which was that its pilots were Lockheed employees on a ferry flight to Puerto Rico. The Air Force and DOD representatives argued that it would be better to use Air Force pilots and state in the event of a mishap that the overflight was a routine Air Force peripheral surveillance mission that had gone off course. McCone then asked Colonel Ledford's opinion of the proposed change. Ledford agreed that the DOD cover story was better but pointed out that the SAC U-2s were much more vulnerable than those of the Agency, which had superior electronic countermeasures and a higher maximum altitude. Ledford then suggested that Air Force pilots use Agency aircraft after receiving familiarization training. After leaving the Special Group meeting, McCone and Gilpatric met with President Kennedy, who approved the San Cristobal mission and the use of Air Force pilots.[14]

Two days later (11 October), Air Force and CIA representatives met to discuss the change in cover stories. Herbert Scoville, CIA Deputy Director for Research, agreed that in the long run the Air Force cover story was best but emphasized that an Air Force pilot should not be used until he had received adequate training. The conversation then turned to the issue of who would run the next mission, the CIA or the Air Force. Strongly favoring Air Force control of the U-2 missions over Cuba, the DOD representatives called DCI McCone and obtained his consent. Shortly thereafter, McCone left

[13] Ibid., p. 31 (~~TS~~ Codeword).

[14] Brig. Gen. Jack C. Ledford, USAF Ret., interview by Gregory W. Pedlow, Washington, DC, 20 February 1987 (~~S~~); Memorandum for DCI McCone from Herbert Scoville, Jr., Deputy Director (Research), "The Chronology of Events Leading to the Transfer of Cuban Overflight Responsibility," 28 February 1963, DCI records, job 80-B-1676R, box 17, folder 18 (~~S~~).

DDCI Marshall S. Carter.

Washington for California and did not return until 14 October. Air Force control of the Cuban overflights became official on 12 October, when President Kennedy transferred "responsibility, to include command and control and operational decisions, with regard to U-2 reconnaissance overflights of Cuba" from the CIA to the Department of Defense.[15] The Air Force then asked to borrow two of CIA's U-2Cs.

The Acting DCI, Lt. Gen. Marshall S. Carter, US Army, reacted strongly to the Air Force takeover of a major CIA operation. At one point he remarked, "I think it's a hell of a way to run a railroad. It's perfectly obviously a geared operation to get SAC in the act."[16] In a series of conversations with high-ranking Air Force and administration officials, Carter argued against changing command and control of the flights at such a crucial time. The Agency operation, Carter pointed out, was already in place and working well, whereas the Air Force lacked experience in controlling U-2 overflights, particularly with the U-2C, which was not in the Air Force inventory. Carter also emphasized that Air Force pilots lacked experience with the more powerful J75 engines in the U-2C. He told Roswell Gilpatric, "To put in a brand new green pilot just because he happens to have on a blue suit and to completely disrupt the command and control and communication and ground support system on 72 hours' notice to me doesn't make a God damn bit of sense, Mr. Secretary."[17] DDCI Carter admitted that the Air Force's cover story was probably better than the CIA's but suggested at one point, "Let's take one of my boys and put him in a blue suit."[18] Realizing, however, that the pilot would probably have to come from the Air Force, Carter concentrated his efforts on trying to convince DOD and administration officials to conduct an orderly transition by allowing the CIA to continue its operation for a few weeks using an Air Force pilot, and the Air Force gradually taking over command and control. Carter's efforts were in vain. The Air Force insisted on immediate control of the operation, and administration officials were unwilling to become involved in what they

[15] Memorandum for DCI McCone from McGeorge Bundy, "Reconnaissance Overflights of Cuba," 12 October 1962, DCI records, job 80-B-1676R, box 17, folder 18 (~~TS~~).

[16] Telephone conversation between DDCI Carter and McGeorge Bundy, 13 October 1962, DCI records, job 80-B-1676R, box 17, folder 18 (~~TS~~ Codeword).

[17] Telephone conversation between DDCI Carter and Roswell Gilpatric, 12 October 1962, DCI records, job 80-B-1676R, box 17, folder 18 (~~TS~~ Codeword).

[18] Telephone conversation between DDCI Carter and Gen. William McKee, 12 October 1962, DCI records, job 80-B-1676R, box 17, folder 18 (~~TS~~ Codeword).

perceived as a jurisdictional dispute. Presidential Assistant for National Security Affairs McGeorge Bundy told DDCI Carter that "the whole thing looks to me like two quarreling children."[19] Furthermore, no one wanted to speak out against a decision that the President had already made.

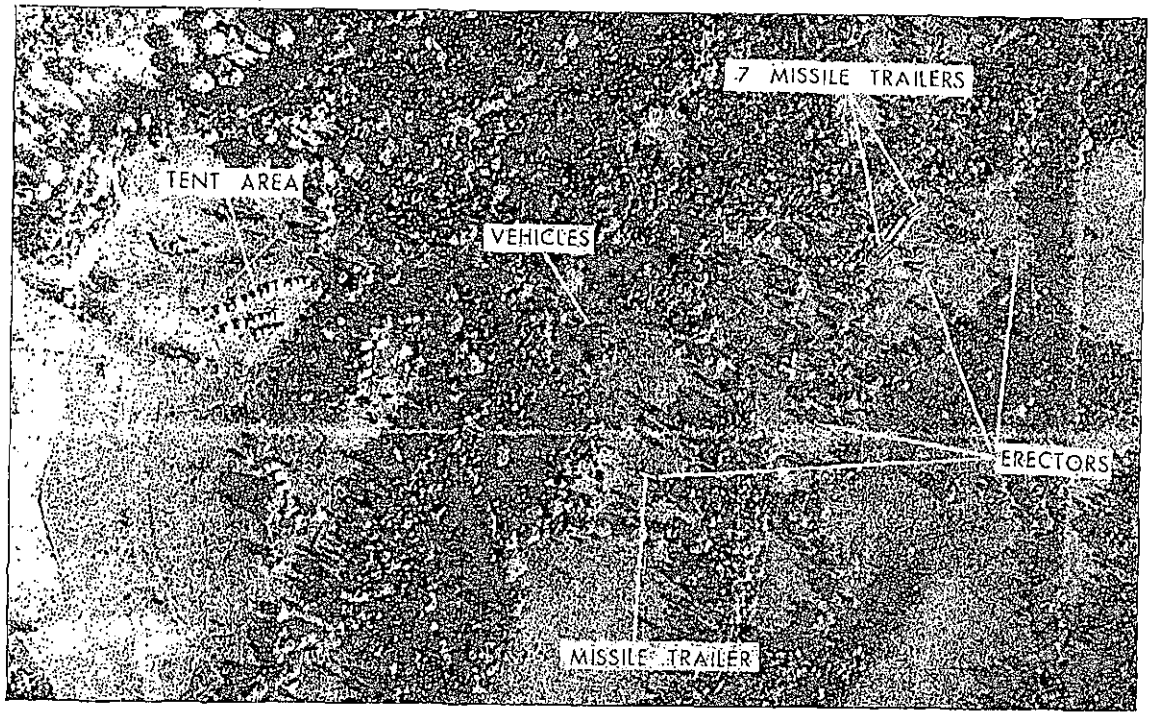
Once the decision was clearly irrevocable, the Agency gave its complete support to the Air Force in preparing for the upcoming overflight. A SAC U-2 pilot had already arrived unannounced at the CIA's U-2 Detachment at Edwards Air Force Base on 11 October, and the CIA U-2 detachment put him through a hasty training program to familiarize him with the U-2C. By Sunday, 14 October 1962, the weather over Cuba had cleared, and the first SAC overflight of the island took place.

When the U-2 returned, its film was rushed to the National Photographic Interpretation Center. By the evening of 15 October, photointerpreters had found evidence of the presence of MRBMs in the San Cristobal area. NPIC Director Arthur Lundahl immediately notified DDI Ray Cline, who in turn notified DDCI Carter (DCI McCone had again left town). As the readout progressed and the evidence became firmer, the DDI notified National Security Adviser Bundy and Roger Hilsman of the Department of State's Bureau of Intelligence and Research, who informed Secretary of State Dean Rusk. On the following morning, 16 October, DDCI Carter briefed the President on the results of the 14 October mission.[20]

Now that the presence of Soviet medium-range surface-to-surface missiles in Cuba had been confirmed, the rules for U-2 mission approval changed. The Strategic Air Command received blanket approval to fly as many missions as needed to cover Cuba completely, without again consulting the Special Group. During the week that followed the discovery of the missiles, SAC U-2s conducted multiple missions each day (see map on page 203). U-2 photography was supplemented by low-level photography taken by high-performance Navy and Air Force aircraft. Throughout the remainder of the Cuban Missile Crisis, the Agency's U-2 pilots remained idle, but the photointerpreters at NPIC did yeoman service in studying the

[19] Telephone conversation between DDCI Carter and McGeorge Bundy, 12 October 1962, DCI records, job 80-B-1676R, box 17, folder 18 (~~TS~~ Codeword).

[20] For a more detailed account of NPIC's discovery of the Soviet missiles in Cuba, see Dino Brugioni, *The Cuban Missile Crisis—Phase I, 29 August–16 October 1962*, DDS&T Historical Series, NPIC-1 (CIA: NPIC, 1971) (~~S~~).



Soviet MRBM site in Cuba, 1 October 1962

thousands of feet of film returned by Air Force and Navy reconnaissance aircraft. President Kennedy used NPIC photographs to illustrate his address to the nation on 22 October 1962, when he revealed the Soviet missile buildup in Cuba and declared his "naval quarantine" to prevent the shipment of offensive weapons to Cuba.

On 27 October, at the height of the crisis, one of the U-2Cs lent by the Agency to the Air Force was shot down over Cuba, killing the pilot, Maj. Rudolph Anderson. This loss again illustrated the U-2's vulnerability to the SA-2 missile. Nevertheless, SAC U-2 overflights continued, both during and after the crisis. Responsibility for photographic coverage of Cuba remained with the Air Force; Agency pilots never flew another mission over the island.

Although SAC carried out most of the U-2 activity during the Cuban Missile Crisis, the Agency's U-2 missions had made vital contributions during the initial stages of the crisis. In all, Project IDEALIST pilots had spent 459 hours overflying Cuba during 1961 and 1962. They had provided concrete evidence of the Soviet buildup on the island, evidence that was simply not available through any

other means. Although by late 1962 photographic satellites had become an integral part of the overhead collection program, only U-2s could provide the highly detailed photography that photointerpreters needed to spot the early stages of work on missile sites. Attempts had been made to photograph Cuba with satellites, but to no avail because the satellites' normal orbits placed them over Cuba at the wrong time of day, after clouds had formed.

U-2s Over South America

Agency U-2s again conducted operations in the Western Hemisphere in December 1963. The Directorate of Plans had requested photographic coverage of Venezuela and neighboring British Guiana because of guerrilla activities conducted by a pro-Castro movement inside Venezuela. Supplies for this movement appeared to be coming across the border from British Guiana. On 30 November 1963, the NSC Special Group approved overflights of the British Guiana–Venezuela border to determine the scope and rate of buildup of guerrilla forces. The Special Group stipulated that the entire effort was to be conducted without the knowledge of either the British or the Venezuelans.

Within three days, several Detachment G aircraft and pilots deployed to Ramey AFB, Puerto Rico, from which they made six flights over the border areas between 3 and 19 December 1963 in an operation known as SEAFOAM. The results of the effort were inconclusive, and the task force returned to Edwards AFB on 22 December.[21]

U-2 OPERATIONS IN ASIA

Detachment C and the Indonesian Revolt of 1958

U-2 operations in Asia began even before the end of overflights of the Soviet Bloc. By 1958 the Eisenhower administration, although very reluctant to approve U-2 flights over or near Soviet and East European borders, was not averse to using the spyplanes in the Third World, where radar detection was unlikely. Thus, in the spring of 1958, Agency U-2s from Detachment C conducted a major reconnaissance effort over Indonesia, Operation ROBIN HOOD.

[21] *OSA History*, chap. 16, pp. 35-36 (~~TK~~ Codeword).

Long unhappy with President Achmed Sukarno's perceived sympathy to Communism and his institution of "guided democracy" in Indonesia, the CIA, after consultation with the State Department, began in early 1957 to supply financial assistance to a group of dissident Indonesian Army officers on the island of Sumatra. By 25 September 1957, the National Security Council had become concerned with the course of events in Indonesia and on its recommendation President Eisenhower authorized the Agency to "employ all feasible covert means" to support the dissidents. Planning for increased aid of all types began immediately, and in January 1958 a US arms shipment for the dissidents arrived in Sumatra. Then on 10 February, the situation came to a head. While Sukarno was out of the country on a state visit to Japan, the dissident army colonels, without consulting CIA, organized a Revolutionary Council in Padang, West Sumatra, and demanded the abolition of President Sukarno's "guided democracy." Five days later, this council proclaimed itself the new "Revolutionary Government" of Indonesia. President Sukarno's armed forces responded swiftly to this threat. In late February the Indonesian Air Force began bombing dissident strongholds, and by mid-March government forces were conducting an all-out air-sea-land drive against the rebel-held areas in central Sumatra. Although the Sumatran rebels were falling back, additional unrest broke out over 1,800 miles away in the islands of Celebes (Sulawesi), and CIA quickly began supplying weapons to these dissidents, too.[22]

Increasingly involved in Indonesia, the Agency urgently needed accurate information on the situation there. As in previous crises, U-2s flew reconnaissance missions. On 24 March 1958, the Development Projects Staff moved the entire complement of Detachment C's pilots and planes from Japan to a base more easily accessible to Indonesia: Cubi Point Naval Air Station in the Philippines. Cubi Point was far from any facility that could develop and interpret the U-2 photographs, so two Photo-Intelligence Division employees went to Clark Airfield, just 30 minutes by air from Cubi Point, to establish a forward processing center. They arrived on 28 March and had the photo lab ready to go on the following day.[23]

Withheld under statutory authority of the Central Intelligence Agency Act of 1949 (50 U.S.C., section 403g)

[22] [redacted], *Covert Support to Indonesian Revolutionary Government, 1957-1958*, 2 vols., Clandestine Service Historical Series, CSHP-53 (CIA: History Staff, 1970) ~~(S)~~; John Prados, *Presidents' Secret Wars: CIA and Pentagon Covert Operations Since World War II* (New York, William Morrow: 1986), pp. 133-144.

[23] *OSA History*, chap. 15, pp. 25-26 ~~(TS~~ Codeword).

Detachment C and the Indonesian Revolt, 1958
China
India
Burma
Laos
VIENTIANE
RANGOON
Thailand
BANGKOK
Cambodia
PHNOM PENH
SAIGON
SOUTH VIETNAM
HANOI
NORTH VIETNAM
Macau (Port.)
VICTORIA
Hong Kong (U.K.)
TAIPEI
Taiwan
East China Sea
Okinawa
TOKYO
Japan
Atsugi
Philippine Sea
Luzon
South China Sea
Foward processing center established, 28 March
Cubi Point Naval Air Station
Clark Airbase
MANILA
Detachment C moved from Japan, 24 March
Philippines
JESSELTON
North Borneo (U.K.)
BRUNEI
Brunei (U.K.)
Sarawak (U.K.)
KUCHING
Borneo
KUALA LUMPUR
Malaya
SINGAPORE
Singapore (U.K.)
Padang
Sumatra
Dissident uprising, 10 February
Government troops retake city, 17 April
US advisors evacuated, 1 May
CIA arms delivery, mid-April
Celebes
CIA support for captured island, 27 April
Morotai
HOLLANDIA
Netherlands New Guinea
Ambon
Capture of Allen Pope, 17 May
Indonesia
DJAKARTA
Java
DILI
Portuguese Timor (Port.)
INDIAN OCEAN
Australia
0 300 600 Kilometers
0 300 600 Miles
~~Secret~~ NOFORN
724755 (R00427) 4-92

The first U-2 mission over Indonesia took place on 28 March 1958. By 12 June, when the operation was phased out, Detachment C U-2s had flown 30 missions over the major islands of Indonesia. Sanitized photos from these missions were used to brief members of the DDP's Covert Action Staff (CAS), who were in charge of a small force of World War II–vintage aircraft such as P-51s and B-26s used to support the rebel troops. The CIA's proprietary, Civil Air Transport, supplied the aircraft, which were based on the Indonesian island of Morotai and flown by mercenary pilots. Desperately short of pilots, the CAS asked if some of the U-2 pilots with experience in World War II aircraft could be detailed to the Morotai effort. Although such a request represented an improper use of the highly trained U-2 pilots and posed a potential threat to the entire U-2 program if one of them were captured, Richard Bissell agreed to send pilots James Cherbonneaux and Carmine Vito to help. Both were experienced with World War II aircraft, although Vito had never flown the rebels' fighter aircraft, the P-51 Mustang. After arriving on Morotai, Cherbonneaux explained to Vito how to fly the fast and powerful Mustang while the two were sitting at a makeshift bar on the edge of the airfield.

Several days later, when Cherbonneaux was off the island on another mission, a flight of Indonesian twin-engine bombers of Czechoslovak manufacture was spotted making its way toward the island. Exclaiming, "I'm not going to sit around and wait to be bombed," Vito had a Filipino mechanic start up a P-51 sitting on the tarmac. In his first and only flight in a P-51, Vito managed to get the plane off the ground. Once he was airborne and turned in the direction of the lumbering bombers, they all took flight in as many directions as there were aircraft. After firing a few .50-caliber rounds in the direction of the closest bomber, Vito circled the field and landed the aircraft safely.[24]

Agency efforts in support of the rebel government proved fruitless. By early May, Central Government forces had taken most of the remaining rebel strongholds, and the Sumatran rebellion was deteriorating into small-scale guerrilla activity. Then on 18 May, an American mercenary pilot, Allen Lawrence Pope, was shot down on a bombing mission over Ambon Island. Pope's capture ended Allen Dulles's enthusiasm for the effort, and President Eisenhower also

[24] Information supplied by Carmine Vito and James Cherbonneaux to Donald E. Welzenbach, May 1986.

wanted no more part of it. The US Government rapidly withdrew its support, and the remaining remnants of the rebellion collapsed. Four years later, the Indonesians freed Pope after Attorney General Robert Kennedy personally appealed to President Sukarno.

When the revolt ended, the U-2s returned to Atsugi. On the way back, one of the planes, which was equipped with a System-V electronic intelligence unit, flew along the coast of China to gather data on Communist Chinese radars.[25]

China Offshore Islands Dispute of 1958

During the summer of 1958, tension between the People's Republic of China and Nationalist China (Taiwan) increased to such an extent that on 18 June Detachment C mounted a U-2 mission to film the Chinese mainland coast and adjacent island areas. On 11 August, People's Liberation Army (PLA) artillery began bombarding the offshore islands of Quemoy and Little Quemoy, where the Nationalists had stationed large numbers of troops to ward off any invasion. On 23 August the Communists increased the shelling. After five days of intense bombardment, which made resupply of the islands from Taiwan impossible, the PLA commander ordered the Nationalist garrisons to surrender, intimating that an invasion was imminent. The Nationalists refused to surrender and received support from the United States in the form of warships from the 7th Fleet, which began escorting Nationalist ships carrying supplies to the beleaguered garrisons.

During this period, Detachment C U-2s flew four missions over the mainland, searching for troop movements that would indicate that the PRC was planning to invade the islands. Photos from these missions showed no evidence of a PRC buildup, but the atmosphere in the region remained tense. Detachment C U-2s flew two more missions (9 September and 22 October) to monitor PRC troop movements and again found no indications of preparations for an invasion. The Offshore Islands Crisis receded in late October 1958 after the PRC learned that it would not receive support from the Soviet Union if the crisis escalated into a confrontation with the United States.[26]

---

[25] Mission folder 1773, (10 June 1958), OSA records, job 67-B-328, box 7 (TS Codeword); *OSA History*, chap. 15, pp. 25-26 (TS Codeword).

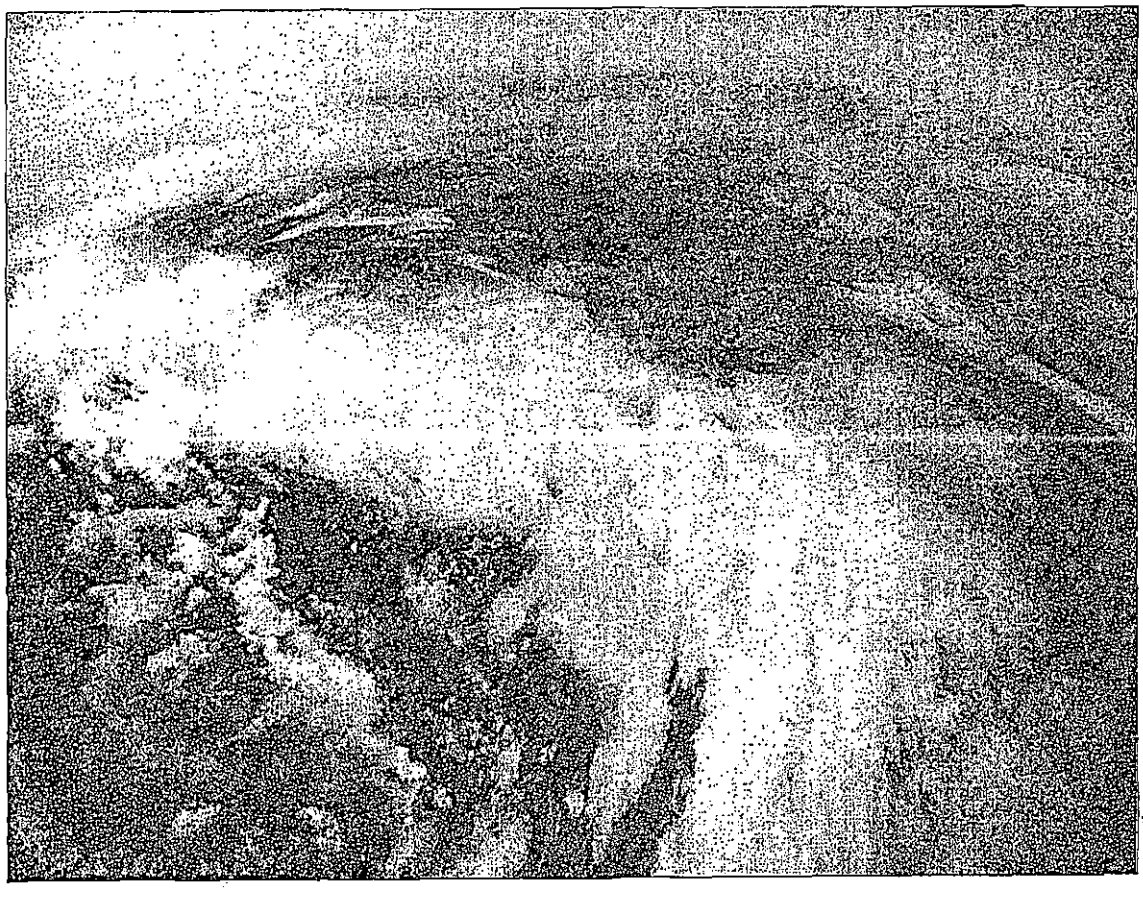
[26] *OSA History*, chap. 15, p. 27 (TS Codeword).

While the Offshore Islands Crisis was still in progress, Detachment C began conducting flights in support of its weather reconnaissance cover story. On 14, 15, and 16 July 1958, U-2s flew high above Typhoon Winnie, which was causing great damage on Taiwan. These missions provided the first photography ever obtained of such a massive storm system. Photographs of the storm were the subject of articles in the magazine *Weatherwise* and the 21 July edition of *Aviation Week*. In September, Detachment C aircraft photographed two more typhoons.

U-2 Support for DDP Operations in Tibet

The consolidation of all Agency air activities under the DDP in 1959 led to increased involvement of the U-2 program with clandestine efforts against Communist governments. One important area of DDP activity during this period was Tibet. In March 1959, the PLA suppressed an uprising against the Chinese occupation of Tibet, and several thousand Tibetans fled the country along with their spiritual leader, the Dalai Lama. Afterward, Agency operatives from the DDP's Far East Division began training some of these Tibetan refugees for paramilitary operations inside Tibet. Once the Tibetans completed their training, FE Division planned to parachute them back into Tibet. Such missions, however, required detailed maps and aerial photographs of the areas of operation. Richard Bissell, therefore, obtained permission from the President to use Detachment C U-2s to provide the necessary photography.

Operation MILL TOWN, as the reconnaissance missions over Tibet were known, consisted of two missions staged from Cubi Point Naval Air Station on 12 and 14 May 1959. The photography revealed that Communist China had built new roads with supply and defense points. Agency photointerpreters also discovered two large new airfields at elevations above 13,000 feet. Later in the year, the Far East Division needed photographs and maps of another area of Tibet. To conceal the target of this new operation, which was codenamed SOUTH GATE, the Development Projects Division planned and flew a total of six missions covering much of Southeast Asia—Vietnam, Laos, Cambodia, and Burma—as well as the desired area in Tibet. Only four of the missions involved the area of operational interest. Five of the flights took place between 29 August and 9 September, and one additional flight (Operation QUICK KICK) followed on 4 November. All of these missions were "Fast Move" operations in which necessary supplies and personnel flew to a remote staging area in a C-130, where they rendezvoused with a U-2 that had been ferried



U-2 photography of Typhoon Winnie, July 1958

in. The staging base in this case was Ta Khli, Thailand. These flights did not go unnoticed; on 13 September 1959, Hong Kong's *China Post* published a story headlined "U-2 of USAF Said Reconnoitering Red China at Unreachable Altitude." [27]

U-2Cs for Detachment C

Late in 1958, Lockheed began refitting the Agency's 13 remaining U-2s with the more powerful Pratt & Whitney J75/P-13 jet engine. The first of these U-2Cs arrived at Detachment C in the summer of 1959. During a test flight of this aircraft (article 360) on 24 September 1959, the pilot decided to set a new altitude record.

[27] Ibid., chap. 18, pp. 6-7, 12; chap. 15, p. 29 (TS Codeword).

Lhasa, Tibet, November 1959

Although the plane was equipped with a camera, it carried no film and did not have a full load of fuel, which made it considerably lighter than an operational U-2C. As a result, the plane reached 76,400 feet—the highest altitude achieved by any of the original U-2 aircraft. In the process, however, the aircraft consumed more fuel than was called for in the test flight plan, causing the engine to flame out during the return to base. The pilot then made an emergency wheels-up landing at a glider-club strip near Fujisawa, south of Atsugi.

The crash did not cause any injuries or serious damage to the aircraft, but it did bring unwanted publicity to the U-2 program. Much of the publicity resulted from the actions of Detachment C's security unit, whose conspicuous Hawaiian shirts and large pistols drew the

attention of Japanese reporters. One reporter even flew over the area in a helicopter, taking pictures of the U-2. These photographs appeared in many Japanese newspapers and magazines.[28]

U-2 Crash in Thailand

Flights by Detachment C U-2s over Tibet and western China continued during the first half of 1960 under Operation TOPPER. The first mission on 30 March was very successful. The second mission on 5 April took good photographs but encountered mechanical problems. At the start of the mission, the landing-gear doors failed to close completely, resulting in increased drag and higher fuel consumption. With no fuel gauge to warn the pilot of the critical fuel situation, the aircraft ran out of fuel far short of Ta Khli, forcing the pilot to make a crash landing in a rice paddy. The area was inaccessible to large vehicles, and the plane, article 349, had to be cut into pieces in order to remove it. With the help of local villagers, the retrieval team dissassembled the aircraft for transport to the base, where the pieces were loaded onto a C-124 under cover of darkness. The crash and subsequent recovery of the U-2 did not attract the attention of the press; there was only one report in a local Thai newspaper, which simply referred to the crash of a jet plane. In appreciation for the assistance provided by the villagers, [redacted] gave the headman funds to build a new school.[29]

25X1, E.O.13526

End of Detachment C Operations

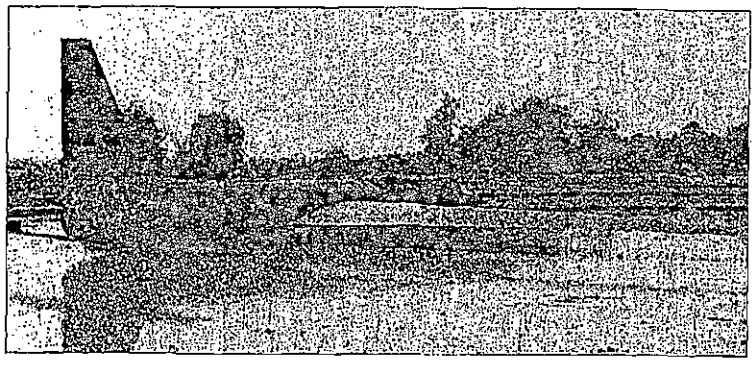
The loss of two aircraft in slightly more than six months left Detachment C with just two aircraft. Fortunately, the level of mission activity remained low because Detachment C was no longer conducting overflights of the Soviet Union.

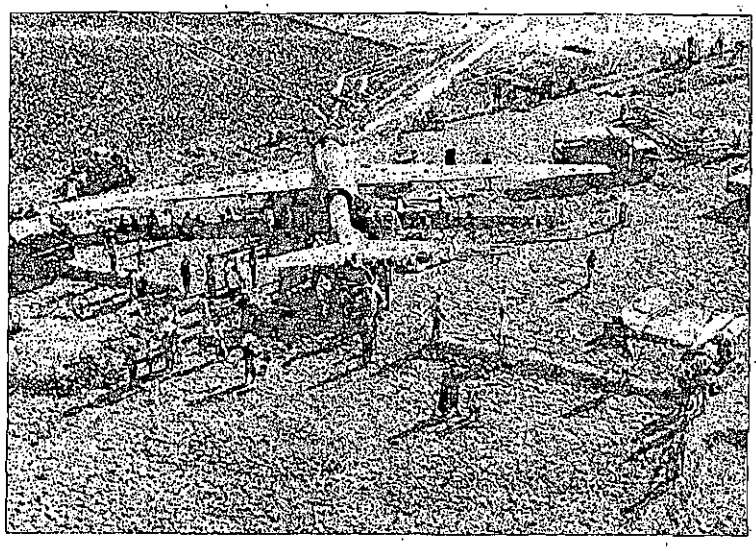
One important remaining mission was high-altitude air sampling (HASP), in which specially equipped U-2s gathered upper-altitude air samples to look for evidence of Soviet nuclear testing. The direction of the prevailing winds made Detachment C ideally situated for this activity, which began in the fall of 1958 and continued in 1959. In late April 1960, Detachment C was preparing to stage to the Philippines to conduct additional air-sampling missions, when the loss of Powers' U-2 temporarily halted all U-2 activities.

---

[28] Ibid., chap. 15, p. 30 (~~CFS~~ Codeword).

[29] Ibid., chap. 15, pp. 32-33 (~~TS~~ Codeword).





Recovery of Article 349, April 1960

The publicity generated by the U-2 incident stirred considerable controversy in Japan, and there were soon demonstrations against the continuing presence of U-2s in Japan. On 6 June 1960, project headquarters decided on a phased-out withdrawal of Detachment C between 15 July and 1 September, but this timetable had to be accelerated when the Japanese Government formally requested the removal of the U-2s on 8 July.[30]

[30] Ibid., chap. 15, pp. 33-36 (~~TS~~ Codeword).

Detachment G Missions Over Laos and North Vietnam

In the aftermath of the Powers loss, both of the overseas U-2 detachments returned to the United States and their aircraft and personnel were incorporated into Detachment G at Edwards Air Force Base in California. This detachment was now responsible for providing coverage in Asia, and its first mission came in Laos. After the neutralist Laotian Government of Souvanna Phouma collapsed in early December 1960, reports began circulating that leftist antigovernment forces were using Soviet arms. Then on 30 December, a new Laotian Government appealed for UN aid against what it said was an invasion from North Vietnam and possibly Communist China. Alarmed over the possibility of the civil war expanding because of the introduction of foreign troops, the Eisenhower administration ordered Detachment G to gather more information on the events in Southeast Asia.

Five Detachment G pilots and planes were ferried to Cubi Point Naval Air Station in the Philippines to conduct an operation known as POLECAT. During the period 3 to 18 January 1961, these U-2s made seven flights over Laos and North Vietnam. To search for the reported foreign troops, these missions concentrated on the lines of communications leading into Laos from North Vietnam and China. In addition, the U-2s scanned North Vietnamese airfields for Soviet aircraft to determine the magnitude of the airdrop operation allegedly supporting the Pathet Lao troops. NPIC sent photointerpreters to Clark Air Force Base in the Philippines to obtain an immediate readout of the results of each mission. The photography did not substantiate the Laotian claims, and on 26 January the Laotian Government retracted its charges of a foreign invasion. Detachment G's U-2s returned to California in early February 1961.[31]

During the final stages of Operation POLECAT, there was a major threat to the security of the mission. The film from the flights made on 16 and 18 January had been sent to the United States for duplicate processing. Afterward the film was put aboard an Agency C-47 on 14 March to ferry it to Washington. During the flight one of the aircraft's engines failed, forcing the crew to jettison 43 boxes of highly classified film over mountainous terrain around Williamsport, Pennsylvania, to keep the craft airborne. After making an emergency

[31] Ibid., chap. 16, p. 17 (~~CPS~~ Codeword).

landing at the Scranton-Wilkes-Barre Airport, the pilot reported the incident to Headquarters. The Office of Security immediately contacted the Pennsylvania State Police, who sealed off the wooded area. Agency security officers soon arrived to search for the boxes. They recovered all 43 containers; not one had broken.[32]

Detachment G's only other activity during the summer of 1961 was a solitary overflight of North Vietnam, known as Operation EBONY. In preparation for this mission, a U-2 deployed to Cubi Point on 13 August 1961. Two days later it successfully conducted the overflight and subsequently returned to the United States.[33]

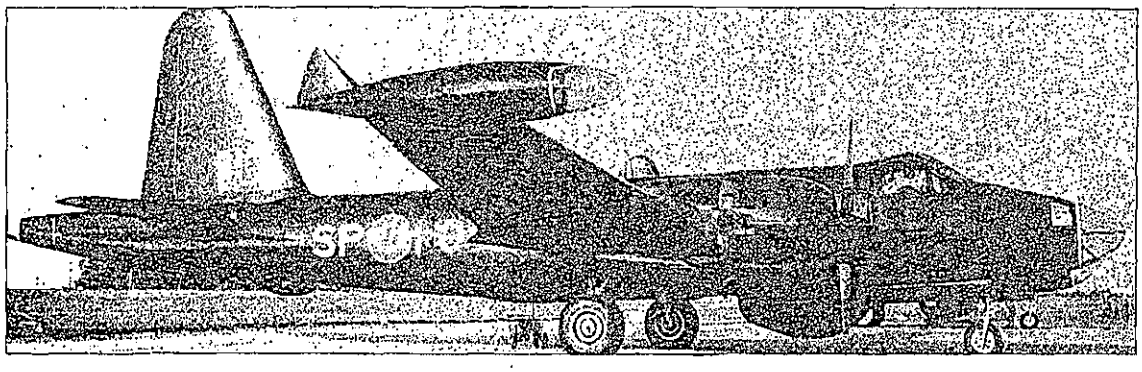
New Detachment on Taiwan

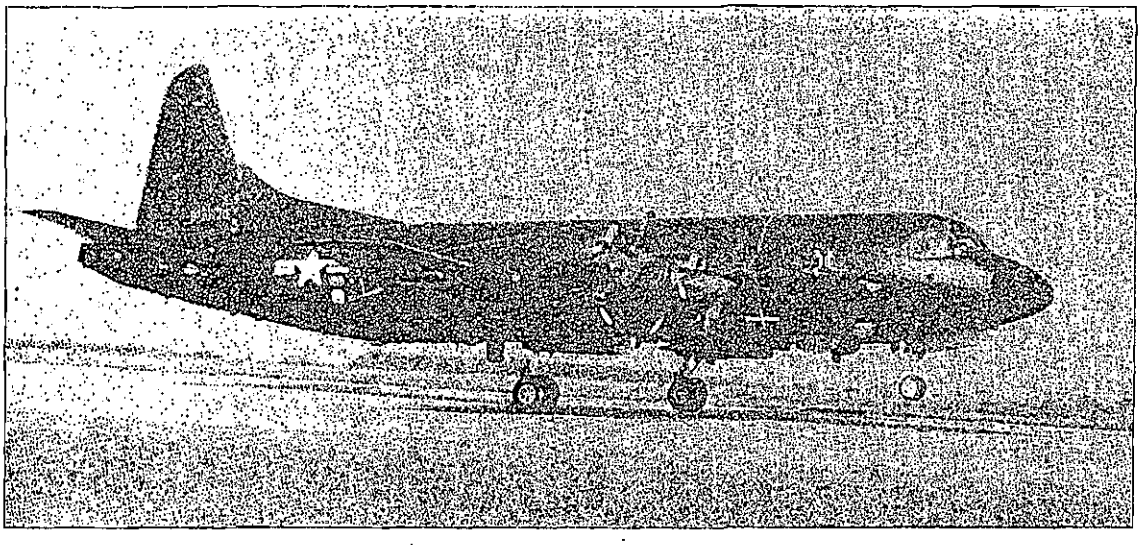
Long before the Nationalist Chinese became involved in the U-2 program, they were flying covert reconnaissance missions for the CIA. In 1952 the CIA began recruiting Nationalist Chinese crews to replace US personnel from the proprietary firm Civil Air Transport, who had been flying Agency aircraft to drop leaflets, agents, and supplies over the Chinese mainland. This project (BGMARQUE) also provided photographic coverage of the rail line from Shanghai to the border with French Indochina. CIA-sponsored aerial reconnaissance over the mainland increased substantially in 1955 with the establishment of Project STPOLLY, which used Agency aircraft with Nationalist Chinese crews to gather Signals Intelligence (SIGINT) and conduct psychological warfare against the People's Republic of China. At first the SIGINT equipment was installed in World War II–vintage aircraft such as PB-4Ys and B-17s, but in 1958 the project received a new aircraft procured covertly by the Agency from Lockheed, the P2V7, with an extremely sophisticated airborne SIGINT system. STPOLLY added the more advanced Lockheed P3A in 1963. Between 1955 and 1967, when the CIA terminated the project, STPOLLY conducted 399 overflights of the People's Republic of China, losing a total of eight aircraft and crews.

In addition to CIA-sponsored aerial reconnaissance projects, the Nationalist Chinese Air Force had its own reconnaissance capability with US-supplied RB-57 aircraft. In 1958 the US Air Force proposed

[32] Ibid., chap. 7, p. 24 (~~TS~~ Codeword).

[33] Ibid., chap. 16, p. 18 (~~TS~~ Codeword).





Lockheed P-2V7 and P3A reconnaissance aircraft

supplying the Nationalist Chinese Government with the most advanced reconnaissance aircraft available, the U-2. The CIA opposed a Nationalist Chinese U-2 program because such flights would destroy the existing unclassified cover for the U-2. In discussions with the Air Force, DDCI Cabell only consented to having Nationalist pilots trained to fly U-2s so that they would be ready in case they were needed in the future; he opposed any Nationalist overflights. The training of the Nationalist Chinese pilots began in March 1959. By the end of the year, there was a group of trained pilots ready for operations, and DCI Dulles met with the Joint Chiefs of Staff to discuss the program's future. Dulles reaffirmed the Agency's opposition to Nationalist Chinese U-2 missions, and the Air Force, which had

wanted the Nationalists to be allowed to begin operations, reluctantly agreed to wait until conditions were favorable.[34] The situation changed radically in May 1960 after the loss of Powers' U-2 destroyed the existing cover story for U-2 operations. Now there was no longer any reason not to use the Nationalist pilots. In addition, the Agency soon found itself in need of a base of operations in the Far East after Detachment C had to leave Japan.

25X1, E.O.13526

During discussions with Nationalist officials on 6 May 1960, the [redacted] raised the possibility of assigning U-2s to the Chinese Air Force. Two weeks later, Gen. Chiang Ching-kuo, head of the National Security Bureau, approached [redacted] informally to propose that the U-2 aircraft based in Japan be moved to Taiwan. This was followed three days later by an official offer on behalf of General Chiang's father, Generalissimo Chiang Kai-shek. President Eisenhower learned of Nationalist China's proposal on 18 June. Several weeks later, Richard Bissell suggested that two U-2s be turned over to the Nationalists for use in overflying the mainland. The project would be conducted along the lines of Project STPOLLY.[35]

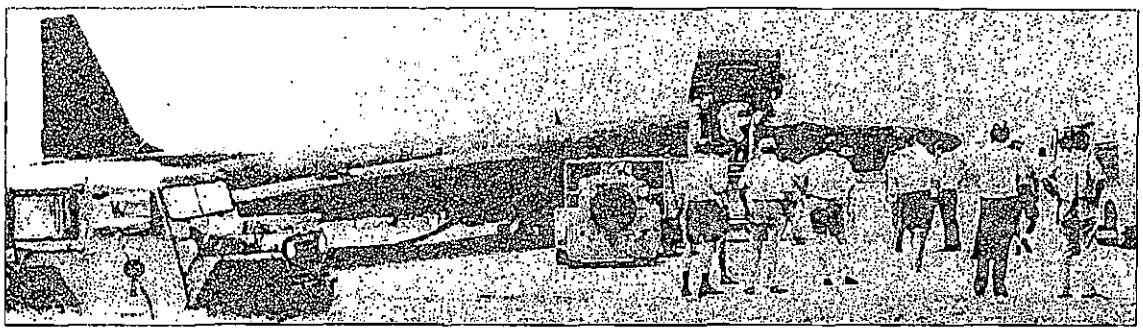
On 26 August 1960, President Eisenhower and the State Department approved Bissell's proposal to turn U-2s over to the Nationalist Chinese rather than move an American detachment to Taiwan. Using Nationalist pilots for overflights had the advantage of providing complete deniability for the United States, even if an aircraft was lost over hostile territory. The U-2s would belong to Nationalist China and would have Nationalist pilots, and there was no overt US involvement with the overflights. In reality, however, the United States would maintain strict control over the missions to be flown.[36]

On 7 December 1960, two U-2s were officially licensed for export to Nationalist China as part of a new effort codenamed TACKLE. These planes came from the Agency's U-2 inventory and arrived in

[34] Andrew J. Russo, *Low-Level Technical Reconnaissance Over Mainland China (1955-1966)* (draft), Clandestine Services Historical Program, CSHP-2.348 (CIA: History Staff, 1972) (~~S~~).

[35] *OSA History*, chap. 17, pp. 1-7 (~~TS~~ Codeword).

[36] Ibid., chap. 17, pp. 12-13, 44 (~~TS~~ Codeword).



Detachment H U-2 at T'ao-yuan Airfield

Taiwan on 14 December. Within the Agency the Nationalist pilots and aircraft were known as Detachment H, and they were based at the Nationalist Chinese Air Force Base at T'ao-yuan. One of the U-2s was painted with the Nationalist Chinese insignia, and the other was left unmarked so that it could also be used by Agency pilots as needed. The planes were maintained by Lockheed mechanics under contract to the CIA. The Agency attempted to maintain at least two U-2s in Detachment H, so lost or damaged aircraft were replaced from the Agency's inventory.

During 1961, Detachment H conducted training missions with both U-2s, and one Nationalist pilot was killed in a crash on 19 March. Although the detachment was ready to begin operations, the new Kennedy administration was not yet ready to authorize overflights of the PRC. In a 3 March 1961 meeting between State Department and CIA officials to discuss the possibility of such overflights, Under Secretary of State Chester Bowles noted that "the President was feeling his way on the international scene, and time was needed to evaluate the new Sino-Soviet posture with relation to the United States." [37] In July 1961 the USIB considered the possibility of conducting overflights of the PRC, but the State Department remained opposed.

By the fall of 1961, interest in overflights of the PRC was growing because of indications that the Chinese were making progress in nuclear energy and missile development. As a result, on 4 October

[37] James A. Cunningham, Jr., Assistant Chief, DPD-DD/P, Memorandum for the Record, "TACKLE STPOLLY Briefing for State Department Officials," 6 March 1961, IC Staff, COMIREX records, job 33-B-119A, box 1, "IDEALIST/TACKLE, 1961" (S).

1961 PFIAB recommended the initiation of a limited number of U-2 photographic missions over the Chinese mainland. The President approved the board's recommendation.

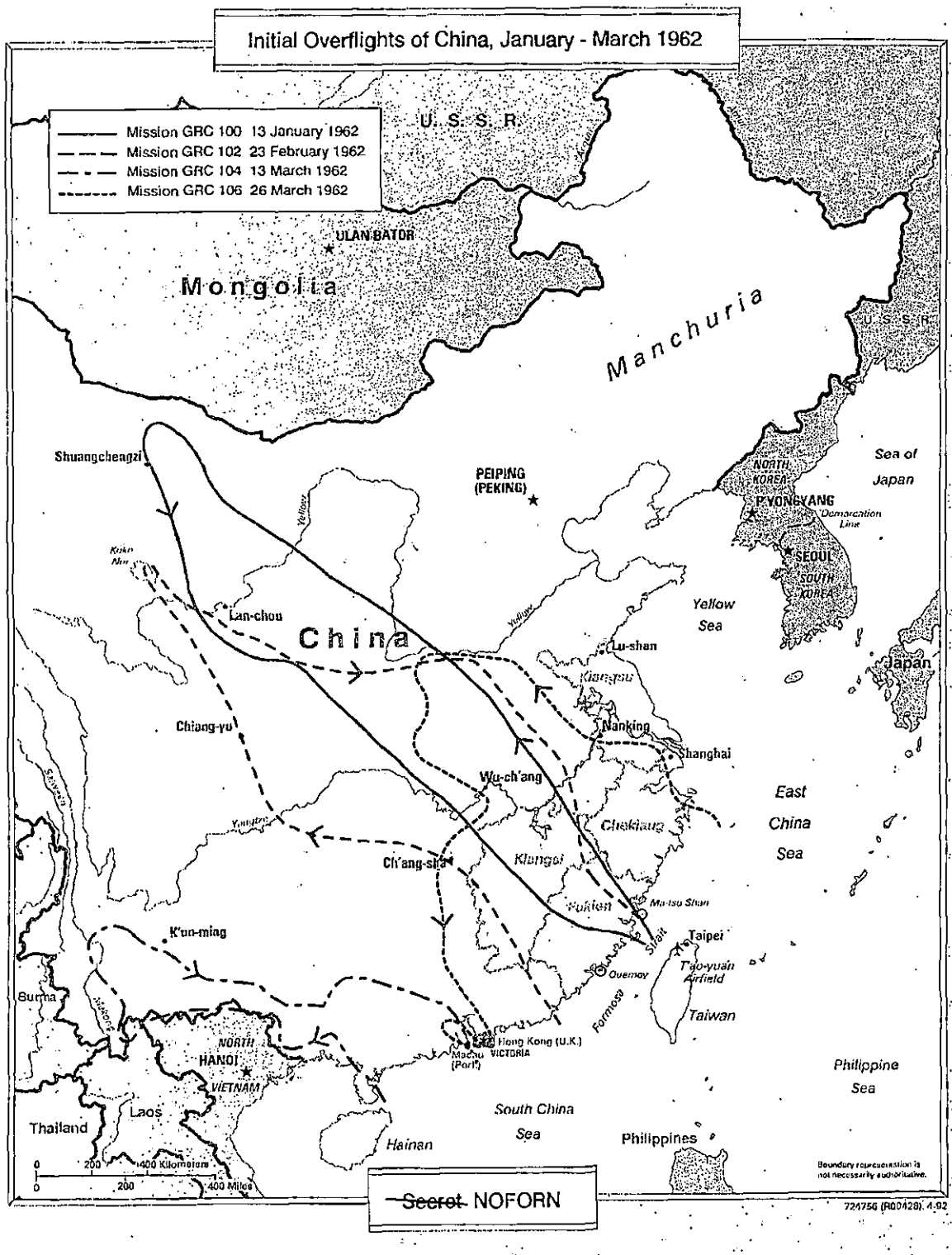
Because the US–Nationalist Chinese overflight program (Project TACKLE) was a joint effort, both countries participated in the approval process and also shared in the results of the missions. The USIB COMOR established the requirements for Detachment H's overflights, which had to be approved by the NSC's Special Group (5412 Committee) and the President. The Nationalist Chinese Government also approved all missions flown by its pilots. Under the terms of an agreement reached with the Nationalist Chinese Government, film from the overflights of the mainland would be processed in the United States, with a duplicate positive copy returned to Nationalist China within 10 days. NPIC was responsible for the initial reporting on these missions.[38]

Project TACKLE overflights began early in 1962. Following a 5 January Special Group decision to approve three missions, a Detachment H U-2 with a Nationalist Chinese pilot flew its first mission over the PRC's missile-testing range at Shuangchengzi on 12 January 1962. Unfortunately, because of faulty navigation or faulty maps, the aircraft was poorly positioned and obtained only oblique, rather than vertical, photography of the range. En route to and from Shuangchengzi, the U-2 overflew Fukien and Chekiang Provinces looking for suspected deployed missiles, but none could be found in the mission photography.[39]

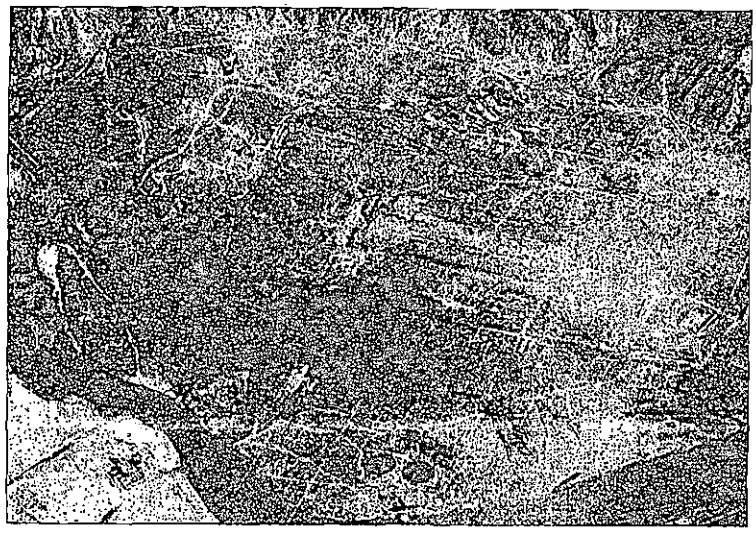
The second Project TACKLE mission took place on 23 February 1962, when a U-2 overflew the PRC's nuclear weapons establishment at Lan-chou. Photography from this mission revealed that the installation was at least two years away from operational capacity. Two more missions on 13 and 26 March flew over K'un-ming and central China covering numerous airfields that had been discovered in satellite photography. The U-2 photographs showed more detail than the satellite pictures, thereby, providing additional intelligence data, particularly

[38] *OSA History*, chap. 17, pp. 18-19, annex 107 (~~TS~~ Codeword).

[39] Ibid., chap. 17, p. 45 (~~TS~~ Codeword); Mission folder GRC100 (12 January 1962), OSA records, job 67-B-972, box 19 (~~TS~~ Codeword). Note: The numbers for missions flown by Nationalist Chinese pilots began with GRC (Government of the Republic of China).



~~Secret~~ NOFORN



Lan-chou, PRC, 23 February 1962

for air order of battle. In addition to the primary targets already described, the initial series of Project TACKLE missions obtained photography of the submarine construction facilities at Shanghai and Wu-ch'ang, which showed a low level of activity. Other photographs revealed tremendous expansion of the industrial complexes at Nanking and Ch'ang-sha and the presence of a previously unknown industrial area at Chiang-yu.[40]

Encouraged by the success of the first TACKLE missions, COMOR recommended in May 1962 that Detachment H cover as many as possible of the highest priority industrial and airfield targets in northeast China and the missile test ranges in north China. COMOR noted that, with the exception of the areas around Peiping and the Shuangchengzi missile test range, the chances of a U-2 being downed were low. The USIB concurred with COMOR's recommendations, and Detachment H therefore conducted three more overflights of the PRC during the month of June.[41]

[40] *OSA History*, chap. 17, p. 45 (~~TS~~ Codeword); Mission folders GRC102 (23 February 1962), GRC104 (13 March 1962), and GRC106 (26 March 1962), OSA records, job 67-B-972, box 19 (~~TS~~ Codeword).

[41] Lay, "USIB History," vol. 2, pp. 385-386 (~~T~~S Codeword).

Before the month was over, however, another confrontation between Nationalist China and the PRC over the Formosa Strait erupted. The Nationalist Government reported a massive buildup of PRC troops and aircraft in Fukien Province opposite the Nationalist-held Quemoy and Ma-tsu Islands. Secretary of Defense Robert McNamara immediately ordered U-2 coverage of the Strait area to determine the extent of the PRC buildup. In response, Detachment H flew six missions over the Strait between 25 June and 28 July 1962. To speed up the readout of this photography, the films were processed at the Asian Photographic Interpretation Center (ASPIC) at Yokota, Japan, a joint military-CIA endeavor. The U-2 coverage ended in late July when it became apparent that the PRC did not intend to mount an invasion of the offshore islands.[42]

The pace of Detachment H missions slowed considerably in August 1962; the sole Project TACKLE overflight covered Peiping and Manchuria. The following month the detachment mounted two missions, one over south China on the eighth and the second over Kiangsu Province on the ninth. Unfortunately, mechanical difficulties led to the loss of the latter aircraft near Lu-shan. A flameout forced the U-2 down to an altitude where PRC interceptors were able to hit the U-2 with an air-to-air rocket. The Nationalist Chinese pilot parachuted and was captured. At this point, President Kennedy ordered a standdown of overflights of the PRC.[43]

Following the capture of the Nationalist Chinese U-2 pilot, the People's Republic of China accused the United States of masterminding the overflights, but the State Department denied any involvement. Nationalist China then revealed that the United States had granted it a license to purchase two U-2 aircraft. In a 13 September 1962 response to the Chinese protest, President Kennedy denied any responsibility for the sale of the U-2s to Taiwan, noting that the sale had occurred under the previous administration. He stated that there were no current plans to sell any more U-2s to Nationalist China. Eight months later, however, the President approved an export license for the delivery of another U-2 to Taiwan. Such licenses were needed only for cover purposes. The Agency continued to maintain two U-2s on Taiwan, bringing new ones in to replace aircraft lost in training or on missions.

---

[42] *OSA History*, chap. 17, p. 46 ~~CIS~~ Codeword).

[43] Ibid., pp. 46-47 ~~CIS~~ Codeword); Mission folders GRC125 (11 August 1962), GRC126 (8 September 1962), and GRC127 (9 September 1962), OSA records, job 67-B-972, boxes 20 and 21 (~~CIS~~ Codeword).

Detachment H resumed overflights of mainland China in December 1962, but its missions now concentrated on the southern portion where there were fewer radars and SAM sites. During December 1962 and January 1963, the detachment conducted two successful overflights of Sichuan, but a mission over south China had to be aborted prematurely. The results of Detachment H's continuing coverage of the People's Republic of China remained of considerable interest to the United States. On 17 December 1962, the Special Group approved plans for fiscal year 1963/64 that included requirements for photo coverage of mainland China and for maintaining at least two operational U-2 aircraft in Detachment H.[44]

Use of Detachment H Aircraft by US Pilots

Detachment H's importance did not lie solely in the missions carried out by its Nationalist Chinese pilots against targets in mainland China; the detachment also provided aircraft for use by American pilots flying missions in other parts of Asia. Indochina was an area of particular interest as American involvement there began growing during the early 1960s. Beginning in February 1962, Detachment G pilots went to T'ao-yuan to use the unmarked Project TACKLE U-2 for overflights of North Vietnam. During the first half of 1962, Detachment G pilots made seven overflights of North Vietnam from the Tao Yuan base. Thereafter, Detachment G pilots could use their own aircraft because the unit began staging teams and aircraft from Edwards AFB to Ta Khli AFB in Thailand.

Between 1962 and 1964, Agency U-2s staged a total of 36 photographic missions over North and South Vietnam. By April 1964, however, photographic requirements were changing from strategic reconnaissance to tactical support as the Viet Cong became more active, taking advantage of the weakness of the South Vietnamese central government following the coup that overthrew President Ngo Dinh Diem in 1963 and subsequent coups by disgruntled army officers. During this period the South Vietnamese "strategic hamlet" concept began breaking down, and the Viet Cong forces stepped up the pace of their attacks. As a result of the increasing level of combat in Indochina, the USIB gave responsibility for aerial reconnaissance of the areas where fighting was taking place to the SAC. Henceforth, SAC U-2s would be used over South Vietnam, parts of Cambodia

[44] *OSA History*, chap. 17, pp. 48-49 (~~TS~~ Codeword); Mission folders GRC134 (25 December 1962), GRC136 (28 December 1962), and GRC138 (20 January 1963), OSA records, job 67-B-972, box 21, and job 66-B-664, box 1 (~~TS~~ Codeword).

within 30 miles of South Vietnam, all of Laos south of Paksane, and all of North Vietnam within 30 miles of South Vietnam or the coast. The remaining portions of Indochina remained the responsibility of the Agency's U-2s. Then in August 1964, following the Gulf of Tonkin Resolution, the Air Force assumed responsibility for all of Indochina.[45]

U-2s in India

In October 1962, the People's Republic of China launched a series of massive surprise attacks against India's frontier forces in the western provinces of Jammu and Kashmir and in the North-East Frontier Agency (NEFA). The Chinese overran all Indian fortifications north of the Brahmaputra Valley before halting their operations.

The Indian Government appealed to the United States for military aid. In the negotiations that followed, it became apparent that Indian claims concerning the extent of the Chinese incursions could not be reliably evaluated. US Ambassador John Kenneth Galbraith, therefore, suggested to the Indian Government that US aerial reconnaissance of the disputed areas would provide both governments with a more accurate picture of the Communist Chinese incursions. On 11 November 1962, Prime Minister Jawaharlal Nehru consented to the proposed operation and gave the United States permission to refuel the reconnaissance aircraft (U-2s) in Indian airspace.[46]

In late November, Detachment G deployed to Ta Khli, Thailand, to carry out the overflights of the Sino-Indian border area. Since the U-2s were not authorized to overfly Burma, they had to reach the target area via the Bay of Bengal and eastern India and, therefore, required midair refueling.

Because of severe winter weather conditions, the first flight did not take place until 5 December. Poor weather and air turbulence hampered the mission, and only 40 percent of the target area could be photographed. A second mission on 10 December was more successful, but the U-2 experienced rough engine performance because of icing of the fuel lines.[47]

---

[45] *OSA History*, chap. 16, pp. 18-19 ~~CPS~~ Codeword).

[46] Ibid., chap. 16, pp. 26-27 ~~CPS~~ Codeword).

[47] Ibid., p. 28 ~~CPS~~ Codeword); Mission folders 3201 (5 December 1962 and 3203 (10 December 1962), OSA records, job 67-B-972, box 26 ~~CPS~~ Codeword).

Detachment G U-2s made four more overflights of the Sino-Indian border areas in January 1963, which led to a PRC protest to India. Photography from these missions was used in January and again in March 1963 to brief Prime Minister Nehru, who then informed the Indian Parliament about Communist Chinese troop movements along the border. Although Nehru did not reveal the source of his intelligence, a UPI wire story surmised that the information had been obtained by U-2s.

The United States had provided photographic coverage of the border area to India for two reasons. First of all, US policymakers wanted a clear picture of the area under dispute. In addition, the intelligence community wanted to establish a precedent for overflights from India, which could lead to obtaining a permanent staging base in India for electronic reconnaissance missions against the Soviet ABM site at Saryshagan and photographic missions against those portions of western China that were out of range of Detachment H. In April 1963, Ambassador Galbraith and the Chief of Station at New Delhi made the first official request to India for a base. The following month, President Kennedy agreed to DCI McCone's suggestion to raise the question of a U-2 base in India when he met with India's President Savepalli Radhakrishnan on 3 June. This meeting resulted in an Indian offer of an abandoned World War II base at Charbatia, south of Calcutta.[48]

The Charbatia base was in poor condition and needed considerable renovation before it could be used for U-2 operations. Work on the base by the Indians took much longer than expected, so Detachment G continued to use Ta Khli when it staged four sorties over Tibet from 29 September to 10 November 1963. In addition to the coverage of the Sino-Indian border during this series of flights, the U-2s also photographed all of Thailand to produce a photomap of the border regions as a quid pro quo for the Thai Government. During one of these photomapping missions, a U-2 pilot conducted the longest mission ever recorded in this aircraft—11 hours and 45 minutes. At the end of this flight on 10 November 1963, the pilot was in such poor physical condition that project managers prohibited the scheduling of future missions longer than 10 hours.[49]

---

[48] *OSA History*, chap. 16, p. 30 (~~TS~~ Codeword).

[49] Mission folder 3238 (10 November 1963), OSA records, job 67-B-972, box 29 (~~TS~~ Codeword).

Charbatia was still not ready in early 1964, so on 31 March 1964 Detachment G staged another mission from Ta Khli. The first mission out of Charbatia did not take place until 24 May 1964. Three days later Prime Minister Nehru died, and further operations were postponed. The pilots and aircraft left Charbatia, but other equipment remained in place to save staging costs. In December 1964, when Sino-Indian tensions increased along the border, Detachment G returned to Charbatia and conducted three highly successful missions, satisfying all of COMOR's requirements for the Sino-Indian border region. By this time, however, Ta Khli had become the main base for Detachment G's Asian operations, and Charbatia served merely as a forward staging base. Charbatia was closed out in July 1967.[50]

Increasing Responsibilities, Inadequate Resources in Asia

The main focus of Agency U-2 activity in Asia remained the U-2s of Detachment H on Taiwan. In March and April 1963, the USIB met to consider COMOR proposals for aerial reconnaissance of Laos, North Vietnam, North Korea, and the People's Republic of China. All of COMOR's intelligence requirements could best be met by the U-2 because heavy cloud cover made it difficult to obtain satellite photography of the region. At the 28 May 1963 meeting of the Special Group, DCI McCone requested authorization for a series of overflights to meet these requirements and stressed the need for additional intelligence on the atomic energy facilities of the PRC. The Special Group then established a "bank" of four authorizations for overflights of the PRC, subject to monthly review by the Group.[51]

As a result of the increasing intelligence community interest in the Far East, both Agency U-2 detachments became very active in the region. Detachment G conducted a number of missions over the border areas of China, North Vietnam, and Laos during April and May of 1963. At the same time, Detachment H became more adventurous, sending U-2s deeper and deeper into the PRC. These missions included renewed overflights of the missile test range near Baotou and the Lan-chou nuclear facilities, as well as targets in northern China, Manchuria, and west-central China (as far as Koko Nor).

---

[50] *OSA History*, chap. 16, pp. 30-34 (~~CPS~~ Codeword).

[51] Lay, "USIB History," vol. 3, pp. 391-392 (~~CPS~~ Codeword).

The increased level of U-2 activity in the Far East during the spring of 1963 exposed a serious weakness in Projects IDEALIST and TACKLE, a shortage of aircraft. The Agency only had seven flyable U-2s when the TACKLE overflights of the PRC began in January 1962, and one of these aircraft had already been lost during an overflight in September 1962. To deal with this shortage, DCI McCone asked Defense Secretary McNamara and the Joint Chiefs of Staff on 10 June 1963 to transfer two U-2s from the Air Force to the CIA. The Defense Department quickly approved this request. Before the two Air Force aircraft were placed in service, however, the Agency had them upgraded with J75/P-13A engines and various electronic devices, a process that took more than four months.[52]

As overflights over the PRC increased, so did concern about the growing number of Chinese surface-to-air missile sites. The Office of Special Activities, therefore, got permission from the Defense Department to equip Project TACKLE aircraft with System-XII SAM-warning units. These devices alerted the pilot that his aircraft was being tracked by the FAN SONG acquisition radar, part of the SA-2's electronic targeting system. The System-XII units also recorded each radar-tracking sequence. Analysis of these recordings revealed changes in the FAN SONG radar's characteristics, information that proved useful in designing electronic-countermeasure (ECM) devices for US aircraft operating over Vietnam during the late 1960s.[53]

Despite the addition of System-XII in the spring of 1963, the Nationalist Chinese–piloted U-2s of Project TACKLE had far fewer ECM devices than other Agency U-2s. Project IDEALIST aircraft possessed a complete suite of ECM gear in addition to the previously mentioned System-XII unit. Among this ECM equipment was a device that told the pilot that an SA-2 missile had been launched (nicknamed the "Oscar-Sierra" unit, which was the acronym for the expletive U-2 pilots used when they learned that an SA-2 missile was on the way: "Oh, shit!") and a System-XIII unit that produced false-angle returns to the homing radar aboard the approaching missile in an effort to steer it away from the aircraft. The Defense Department opposed installing such devices aboard Detachment H's U-2s, for fear they could fall into Communist hands.

---

[52] *OSA History*, chap. 16, p. 10 (~~TK~~ Codeword).

[53] Ibid., chap. 17, p. 50 (~~TK~~ Codeword).

The danger posed by the growing number of SA-2 sites in the PRC was clearly demonstrated on 1 November 1963, when a second Project TACKLE U-2 was lost near the Kiangsi-Chekiang border on its way back from photographing the PRC's Shuangchengzi missile test range. As was the case after the first operational loss over China in September 1961, President Kennedy ordered a standdown of overflights of mainland China. This standdown lasted almost five months.

As a result of this second loss over the PRC, the Office of Special Activities began installing a new 30-channel telemetry system aboard Detachment H U-2s to monitor various aircraft functions. Known as BIRDWATCHER, this unit periodically broadcast a burst of data to the airbase that launched the U-2. This data burst contained a status report on all the major systems aboard the plane, such as airspeed, altitude, exhaust temperature, fuel supply, film supply, and oxygen supply. BIRDWATCHER provided project managers with a benchmark of aircraft performance that could be used to determine if a lost plane had been shot down at altitude or had suffered mechanical failure.[54]

BIRDWATCHER's first operational use came on 16 March 1964, when overflights resumed with a mission over southern China. The PRC was now a high-priority target for the U-2 because more data were needed to prepare National Intelligence Estimates due in the autumn. Of particular concern was the PRC's nuclear program. Despite the high priority of its missions, Detachment H's resources remained scarce. It was short of both pilots and planes and never had more than three U-2s or six qualified Nationalist Chinese pilots at any one time. By the spring of 1964, crashes during training and the two losses over the mainland had reduced Detachment H to only two qualified pilots, one of whom suffered from ulcers and a nervous disorder. Indeed, this pilot had every reason to be nervous; he flew three of the next four Detachment H overflights and became the third Nationalist Chinese pilot to be shot down over the mainland.

This loss came on 7 July 1964. The Nationalist Chinese pilot's last transmission was that his System-XII unit had alerted him that he was being tracked by the FAN SONG radar. BIRDWATCHER data revealed that the aircraft was at penetration altitude and all systems

[54] Ibid., p. 51, 53 (~~CPS~~ Codeword).

were normal when the pilot made this report. Project managers presumed that the U-2 was downed by a direct hit or near miss by an SA-2 missile.[55]

President Johnson ordered a standdown of overflights of the PRC. This standdown was welcomed by the Nationalist Chinese Government, which told the Taiwan Chief of Station that it wanted "to let some time go by" before more overflights were scheduled. The Nationalists pointed out that the only remaining qualified U-2 pilot had "disqualified" himself because of nervous tension. No new pilots could be qualified for U-2 flights before mid-August.

The Nationalists then demanded faster and higher flying aircraft as well as better antimissile equipment for the planes. This request led some CIA personnel to suspect that Nationalist China had learned about Project OXCART, the successor to the U-2 that was still undergoing testing. Despite the Nationalists' request for better ECM equipment, the Defense Department remained reluctant to authorize the use of the System-XIII false-angle radar jammer on Project TACKLE U-2s. The Defense Department feared that the loss of this device with its highly advanced traveling-wave tube (TWT) would enable Communist Bloc technicians to devise countermeasures and also learn how to produce the highly efficient TWT themselves. As an incentive for the Nationalist Chinese to agree to more overflights, the CIA agreed to permit them to process the U-2 film on Taiwan and to use their own photointerpreters to exploit the film along with US photointerpreters.

To counter the shortage of pilots in Detachment H, DCI McCone suggested to the Special Group on 6 August 1964 that civilian CIA pilots be used to fly missions over mainland China. The group agreed that the matter should be taken up with President Johnson. On the following day, however, Presidential National Security Assistant McGeorge Bundy informed McCone that, because Secretary of State Rusk and Secretary of Defense McNamara opposed the idea, he would not take it up with the President.[56]

---

[55] *OSA History*, chap. 17, pp. 53-55 (~~TS~~ Codeword); Mission folder C174C (7 July 1964), OSA records, job 66-B-664, box 7 (~~TS~~ Codeword). Note: beginning in 1964 Nationalist Chinese mission numbers began with "C" instead of "GRC."

[56] *OSA History*, chap. 17, pp. 58-59 (~~TS~~ Codeword).

Advanced ECM Equipment for Detachment H

Demand for overhead photography of the PRC continued to grow, spurred in part by the results of earlier U-2 missions that revealed the presence of Soviet-made MiG-21s in the PRC. In addition, there were indications that Communist China might be producing its own SAMs. Furthermore, satellite photography revealed that preparations for the first Chinese nuclear test were almost complete at the Lop Nor test site.

The need for photographs of the Lop Nor site was considered so urgent that the Defense Department finally relented and permitted the System-XIII false-angle device jammer to be installed in Project TACKLE aircraft, with the proviso that it not be turned on until after the pilot had been alerted by System-XII that he was being tracked by FAN SONG radars. Photographing Lop Nor, however, was not a simple task. Located more than 2,000 miles west-northwest of Taiwan, Lop Nor lay beyond the round trip range of T'ao-yuan–based U-2s and in-flight refueling was not possible. Lop Nor was closer to Ta Khli, Thailand, only 1,650 miles northwest of that base, and much closer to Charbatia, India, which lay only 1,200 miles south of the testing site.

After refusing DCI McCone's suggestion to stage a Lop Nor overflight from Charbatia using a CIA civilian pilot, President Johnson approved a proposal to send a Project TACKLE unit to Ta Khli for the mission to Lop Nor. A Detachment H U-2 with a Nationalist Chinese pilot deployed to Ta Khli in mid-October to prepare for the overflight. Before mission preparations could be completed, however, the Chinese detonated their first nuclear weapon on 16 October 1964, and the mission was canceled.[57]

The first overflight of Communist China since the 7 July 1964 loss was a 31 October mission over Lan-chou. By mid-November, three more overflights had taken place, one over North Korea and northern China and two over southern and central China. An overflight of Manchuria on 9 December 1964 brought back photos of a G-class ballistic-missile submarine. By this time, the mass of data being provided by Project TACKLE overflights was overwhelming the analysts of the Office of Research and Reports (ORR). An ORR memorandum from 11 January 1965 reported that analysis of the "large backlog of unexploited photography on Chinese Communist

[57] Lay, "USIB History," vol. 6, pp. 751, 753-755.

ground force installations...would require about two man-years work, backed up by a larger expansion of photointerpretation effort." [58]

Use of Infrared Scanner Over PRC Nuclear Plants

Photographic missions were not the only method used by the Agency to gather information on the PRC's progress in developing nuclear weapons. In fact, standard overhead photography was of little value in determining whether or not various nuclear-related facilities were producing weapon-grade materials. Such information could only be obtained from agent reports or infrared imaging of the targets, which would reveal the heat associated with nuclear-refining activity. Early in 1963 the Agency sought special funds from the Pentagon to develop a high-resolution infrared scanner for the U-2 in order to ascertain the operational status of Chinese nuclear reactors. After this request was turned down, the Agency decided to fund the development of such a scanner itself. In the spring of 1963, the Office of Research and Development (ORD) signed a contract with the Texas Instruments Corporation for the production of an infrared scanner that would operate in the 0.5-milliradian range. To conceal the intended purpose of the scanner, the devices were called forest fire detectors and assigned model numbers such as "FFD-1."

Tests conducted in September 1963 proved that model FFD-2 was capable of revealing activity within US nuclear-processing facilities at Oak Ridge, Tennessee, and Hanford, Washington. Further improvements followed, and on 26 October 1964 ORD declared the infrared unit operational. In reporting the successful results of the infrared testing to DCI McCone, ORD Director Robert Chapman mentioned that there was only one such unit in existence. McCone immediately told Chapman to order more, and the Agency then purchased several for Project TACKLE and a stereo-configured FFD-4 for Project OXCART.

DCI McCone reviewed these infrared detection developments with Defense Secretary McNamara and his new deputy, Cyrus Vance, on 30 October, being careful to point out that the new device was a product of Agency-sponsored R&D because the Pentagon had refused to fund it. He suggested that the scanner be flown over Lop Nor, Baotou, Youmen, and Lan-chou in U-2s with Agency pilots staged

[58] ORR, Memorandum for the Record, 11 January 1965, OSA records (~~CRS~~ Codeword).

from Taiwan and Charbatia. As in the past, the administration opposed the use of Agency pilots for overflights of China, so the device was sent to Detachment H on 7 November 1964.

The FFD-2 unit was installed in a U-2C in order to gather data from a suspected uranium gaseous-diffusion plant at Lan-chou. It took four tries, however, before this goal was achieved. The first attempt, which was scheduled to be staged from Kunsan, South Korea, on 22 November 1964, had to be abandoned after the pilot became ill. A second mission took off three days later with another pilot. Flying at 69,000 feet, the plane was tracked by radar during the entire flight and attracted a total of five PRC fighters. When the U-2 was within 10 miles of Lan-chou, a previously unknown SA-2 battery launched several missiles. This unexpected attack so unnerved the pilot that he broke off the mission without overflying the target area. Three weeks later a third mission to Lan-chou was aborted because of bad weather shortly after the U-2 entered PRC airspace. The pilot did not bother to return to Kunsan but flew directly to T'ao-yuan. Finally, on 8 January 1965, a Detachment H U-2 made it all the way to Lan-chou and back to Kunsan with the infrared scanner and obtained good results. When the film was processed and evaluated at NPIC, analysts determined that the gaseous-diffusion plant was indeed operating.[59]

Two days after the successful Lan-chou mission, the infrared unit was airborne again, this time bound for Baotou and the air-cooled reactor located nearby. This flight ended abruptly when an SA-2 missile downed the U-2 26 miles short of the target. It was the fourth U-2 lost over China. No more U-2 flights over China used the infrared scanner, although a later model IR device was shipped to Tao Yuan and kept in readiness in the event it might be needed.

The loss of yet another U-2 and its pilot made Nationalist Chinese officials reluctant to resume overflights of the mainland. They insisted that their U-2 pilots be given permission to turn on the System-XIII false-angle device during the entire time they were over hostile territory, not just after they had been alerted by the System-XII radar-detection device. The Defense Department acceded to these demands in order to keep the TACKLE flights going. In addition, the Project TACKLE U-2s began receiving even more advanced ECM equipment.[60]

---

[59] Mission folders C284C (22 November 1964), C304C (25 November 1964), C344C (19 December 1964), and C015C (8 January 1965), OSA records, job 66-B-664, box 8, job 69-B-404, box 2, and job 66-B-597, box 1; *USIB History*, vol. 6, p. 758 (~~TS~~ Codeword).

[60] Ibid., chap. 17, pp. 59-61 (~~TS~~ Codeword).

With their demands met, Nationalist Chinese officials again consented to overflights by Detachment H, and operations resumed in February 1965 with three missions over the mainland. By this time US interest in the People's Republic of China was very high because of the PRC's development of nuclear weapons. The Special Group, therefore, approved an extensive reconnaissance program directed against the PRC. By the end of the year, Detachment H had flown 30 missions, the highest annual total during the entire program.

The level of activity declined during 1966, with only 10 missions flown over the mainland. Detachment H also suffered the loss of two more aircraft and pilots in crashes during training missions in 1966. In the fall of that year, joint US–Nationalist Chinese relations in the field of overhead reconnaissance were further strained by the unilateral US decision to kill the longstanding program of low-altitude nighttime overflights of the mainland (STPOLLY).[61]

25X1, E.O.13526

[61] Ibid., chap. 19, annex 120, pp. 22-24; chap. 17, p. 69 (~~TS~~ Codeword).

25X1, E.O.13526

25X1, E.O.13526

The entire U-2 overflight program was temporarily halted in early November 1967 after an Air Force U-2 in Vietnam was discovered to have cracks in its wing. All Air Force and CIA U-2s were ordered back to Lockheed for ultrasonic inspection of the wings and other stress points to check on metal fatigue. Upon completion of this inspection, both the Air Force and the Agency resumed their overflight activity.[64]

The End of U-2 Overflights of Mainland China

Project TACKLE mounted a mission over northeastern China on 13 December 1967 and an overflight of central China on 5 January 1968. Later that month the Far East became very tense when North Korea seized the US electronic intelligence ship Pueblo on 23 January. One

[63] Ibid., chap. 17, p. 72 (~~TS~~ Codeword).

[64] Mission folder C297C (8 September 1967), OSA records (~~TS~~ Codeword).

[65] *OSA History*, chap. 16, p. 44 (~~TS~~ Codeword).

week later the Viet Cong and North Vietnamese launched their Tet offensive in South Vietnam. The 303 Committee (the new name for the Special Group after 1964) decided on 1 February 1968 to suspend a group of overflights scheduled for February and called for mission-by-mission approval "during this period of tension." The committee approved one additional overflight of southern China, which was flown by Detachment H on 16 March 1968, and two overflights of Cambodia, carried out on 27 March and 3 April 1968 by Detachment G in its first operations since early 1966. These three missions turned out to be the last overflights by U-2s in the Far East. By this time U-2 flights over the PRC had become so dangerous that the State Department opposed further overflights, and on 10 April 1968 the 303 Committee decided not to approve any mission that would fly closer than 20 miles from the coast of China.

One reason why Detachment H's overflights were stopped was the steady increase in the PRC's ability to track and engage U-2s, as evidenced by its success in downing five U-2s. By 1968 PRC radars along the coast opposite Taiwan were keeping a close watch on U-2 activity from the T'ao-yuan base and actively tracked U-2s as soon as they became airborne. The U-2s then had to face a growing PRC air defense system that not only consisted of SA-2 missiles but also the fast and high-flying MiG-21. The PRC's MiG-21 pilots had become adept at the power-zoom technique and were threatening almost every U-2 mission. The risks to U-2s now seemed too great.[65]

The decision to end Asian overflights was also rooted in the Johnson administration's change in its whole approach to the war in Indochina in the spring of 1968. On 31 March 1968, the President limited the bombing of North Vietnam in order to improve the chances for peace talks. The end of flights over the People's Republic of China was viewed as another way to improve the peace process.

During its six years of overflight operations, Project TACKLE had been extremely active, staging a total of 104 flights over the People's Republic of China (see table, page 244). These missions had brought back huge amounts of data about the PRC but not without a price: five Project TACKLE aircraft had been downed by the PRC, with two of the pilots killed and the other three captured (although later returned). In addition, five other Nationalist Chinese pilots had been killed in training accidents.

[65] OSA, "Chronology of Events Leading to Present Impasse on Resumption of TACKLE Overflights of Mainland China," December 1968, OSA records ~~(TS)~~.

Overflights by Project TACKLE

Fiscal Year	Overflights
Total	**104**
1962	18
1963	17
1964	13
1965	30
1966	10
1967	14
1968	2

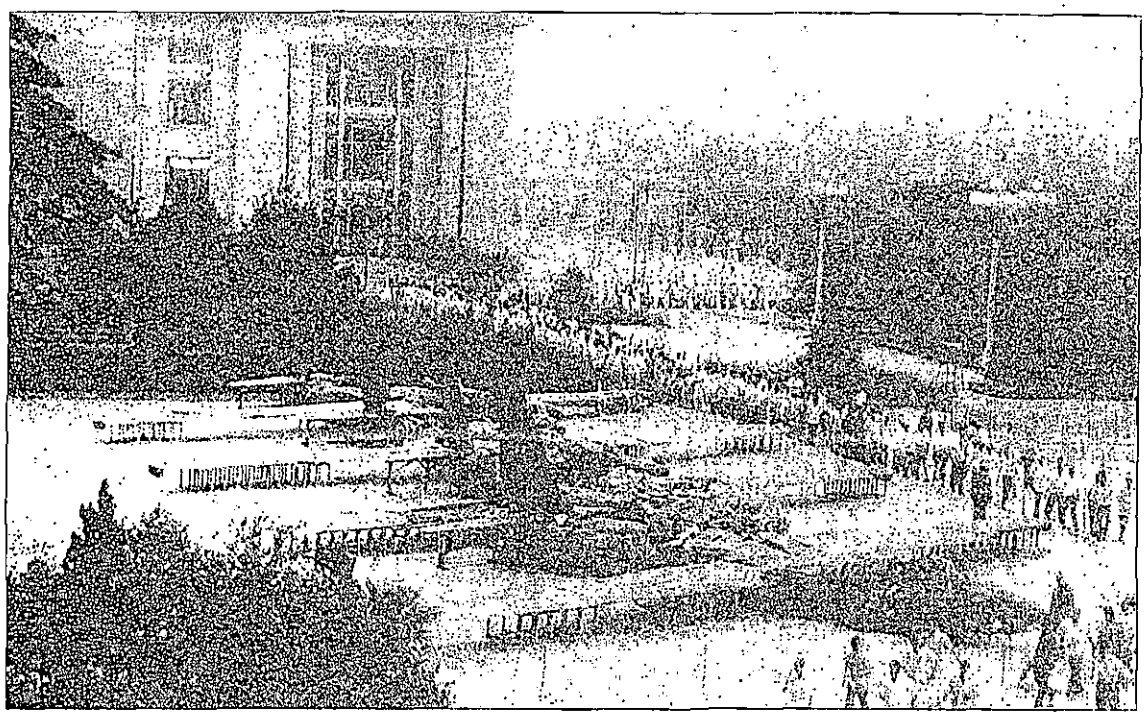
Peripheral Missions by Detachment H

Detachment H did not cease its activities following the termination of overflights of mainland China. Its next U-2 mission took place on 18 May 1968. This was an electronic intelligence mission that, in accordance with the new guidelines, never came closer than 20 miles to the Chinese coast. All future Detachment H missions against the PRC also conformed with this restriction but were still the target of interception attempts by PRC MiG-21s or hastily erected SAM sites on offshore islands. The use of peripheral missions prevented any further losses, although one aircraft crashed into the sea from unknown causes shortly after taking off to start a mission on 5 January 1969. Another pilot was killed on 24 November 1970 in a crash during a routine training mission.[66]

The level of activity gradually increased during the remaining years of Project TACKLE, as can be seen in the table on page 245.

The U-2s flying these missions were generally equipped with the B or the newly designed 48-inch H cameras to obtain oblique photography and with various signal-intelligence-gathering systems. Beginning in May 1971, Project TACKLE began using a new communications-intelligence collection package known as LONG SHAFT. This system was used on 32 occasions through 12 December 1973, when the LONG SHAFT collection program ended.

[66] "Office of Special Activities History, April 1969 to Phase-Out," (draft; hereafter cited as "OSA History-2") (CIA: DS&T, 1974), chap. 3, pp. 36-42 (~~TS~~ Codeword).

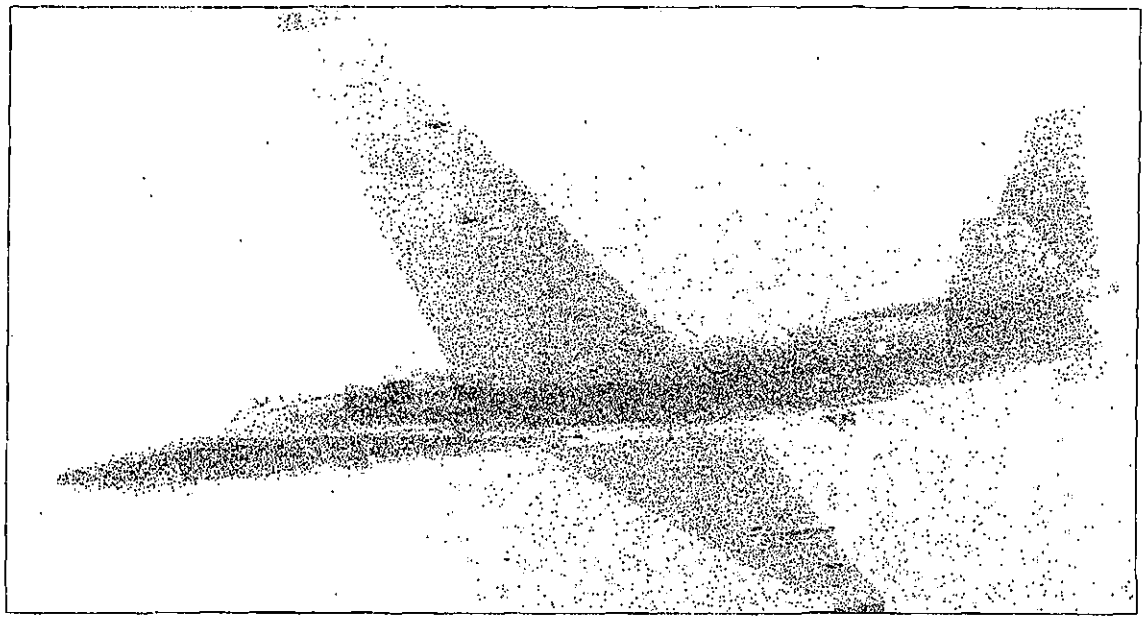


Display of downed detachment H U-2s in Peiping

Project TACKLE Peripheral Missions, 1969-1974

Fiscal Year	Missions
1969	9
1970	14
1971	19
1972	23
1973	31
1974	17

Once the United States began seeking a rapprochement with the People's Republic of China, Detachment H U-2s came under more and more restrictions. Soon after the impending visit of President Richard M. Nixon to the PRC was announced, U-2 missions were ordered to stay even farther away from the mainland: 25 nautical miles instead of the previous 20. During the months of February and



Nationalist Chinese U-2R

March 1972, when the President's visit took place, Detachment H ceased all operational missions.[67]

In March 1973, the TACKLE agreement with the Nationalist Chinese was renegotiated. Although no end date was set, the agreement contained a termination clause that would become effective three months after notification by either party. This clause provided more flexibility to the United States, which could now end the Nationalist Chinese U-2 program whenever US foreign policy considerations made such a step desirable.

Operation SCOPE SHIELD Over North Vietnam

In addition to the Project TACKLE peripheral missions against the PRC, Detachment H (with Agency rather than Nationalist Chinese pilots) flew a series of missions known as Operation SCOPE SHIELD to gather intelligence on activities in North Vietnam. The Indochina area had become the responsiblity of the Air Force in 1964, but, under the terms of the cease-fire agreement negotiated with North

[67] Ibid., pp. 44-45 (~~TS~~ Codeword).

Vietnam in January 1973, US military flights in the area were forbidden. The Nixon administration, therefore, tasked the CIA with monitoring North Vietnam's compliance with the cease-fire accords.

The Agency dispatched several pilots to Taiwan under the cover of Lockheed employees working on a government contract to check weather conditions. Their highly sensitive missions had to remain at least 15 nautical miles away from the North Vietnamese coast, and they initially flew at low altitude in a deceptive direction in order to avoid PRC radars. These constraints made the missions difficult because at low altitude the U-2 consumed more fuel and encountered more turbulence and the pilots' pressure suits tended to overheat.

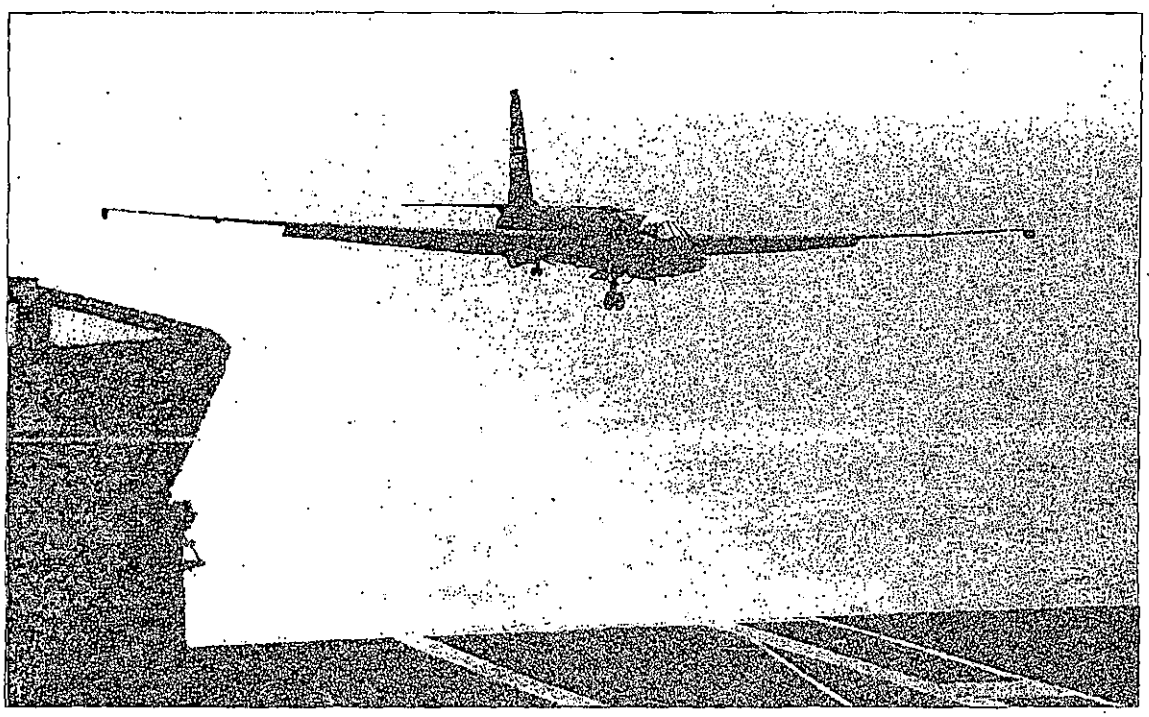
The first mission on 30 March 1973 was only marginally successful because of cloud cover and haze, which prevented it from photographing most of its targets. A second mission on the following day had somewhat better luck with the weather, but problems with the film processing reduced the mission's coverage. Afterward, the monsoon season prevented any further missions until 21 July 1973. This mission obtained usable photography of SAM sites and North Vietnamese supply operations, although the resolution was not as high as it should have been because the H camera lens had not been properly focused. The last SCOPE SHIELD mission, on 6 January 1974, finally succeeded in obtaining high-quality photography. The mission provided complete coverage of shipping in Haiphong Harbor, SAM defenses, and North Vietnamese naval order of battle.[68]

IMPROVEMENTS IN U-2 TECHNOLOGY

Modification of U-2s for Aircraft Carrier Deployment

In mid-1963, the Office of Special Activities set in motion Project WHALE TALE to examine the possibility of adapting the U-2 aircraft for operations from an aircraft carrier. In the past, protest notes from the Soviet Union to Turkey and Pakistan and from Communist China to India had been responsible for interrupting overflight operations. CIA planners believed that, if U-2s could be modified to operate from aircraft carriers, the United States could avoid the political problems

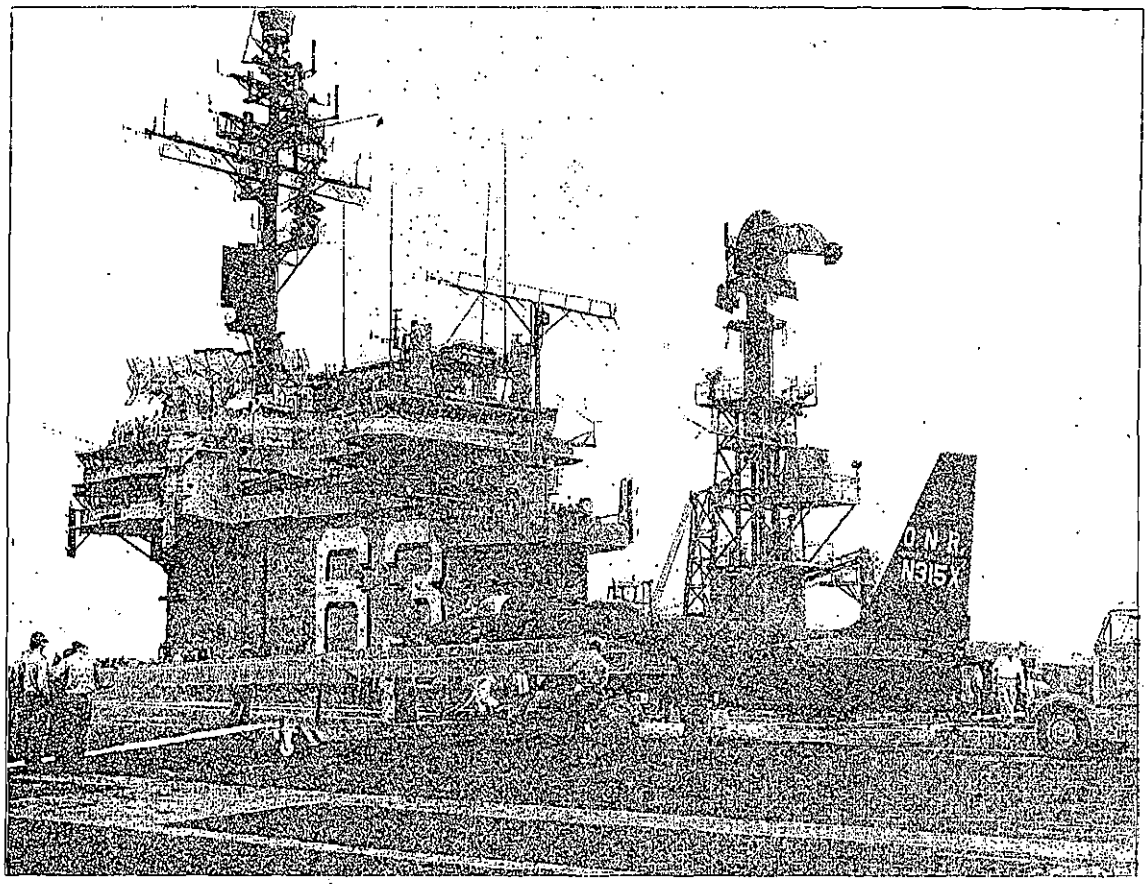
[68] Ibid., pp. 48-51 (~~CRS~~ Codeword).



U-2 on the USS Kitty Hawk, 5 August 1963

involved in seeking permission to base U-2s in other nations. Kelly Johnson began working on changes to the aircraft, and Office of Special Activities Deputy Director James A. Cunningham, Jr., a former Marine Corps aviator, asked the Navy for assistance.

The first test of the U-2's capability for carrier operations took place in August 1963 from the USS Kitty Hawk operating in the Pacific Ocean off San Diego, California. A U-2C, which had been loaded aboard the carrier at North Island Naval Base, took off from the flight deck with a full load of fuel and was airborne within 321 feet. No assistance from catapults was necessary. Although the takeoff was very successful, the attempted landing was not. The aircraft bounced, hit hard on one wing tip, and then just barely managed to become airborne again before reaching the end of the deck. Kelly Johnson realized that the airframe would have to be altered in order to make carrier landings possible. These alterations involved strengthening the landing gear, installing an arresting hook at the rear of the fuselage, and fitting "spoilers" on the wings to cancel the aerodynamic lift once the aircraft was over the flight deck. Aircraft thus modified were designated U-2G. While several aircraft



U-2 on the USS Kitty Hawk, 5 August 1963

underwent these modifications, Detachment G pilots began undergoing training in landing on aircraft carriers. The first successful carrier landing took place on 2 March 1964.[69]

Use of Carrier-Based U-2 To Film a French Nuclear Test Site

Within a few months after the completion of carrier testing, one of the carrier-modified U-2s conducted an operation in the Pacific. Its mission was to gather information on the activities of an ally. In

[69] "U-2 Aircraft Carrier Operations: Project 'WHALE TALE,' Operation FISH HAWK," DS&T, 1964, pp. 1-13 (~~TK~~ Codeword); Johnson, "Log for Project X," 5 August 1963 and 2 March 1964.

December 1963, France had announced its intention to detonate a hydrogen device over Mururoa Atoll in the Tuamotu Archipelago area of French Polynesia but had given no specific date for the event. The Committee on Overhead Reconnaissance had been following French nuclear developments since September 1963, when it had apprised the USIB of the need for overflights of this South Pacific area. At that time the USIB decided against recommending such overflights because of State Department concern about potential political difficulties with France in the event the mission was discovered.

Following reports of a buildup of French troops and technical personnel in neighboring Tahiti, the Special Group on 24 April 1964 approved a mission to overfly the atoll to check for activity. This required photography with a resolution better than the 3 to 5 feet possible with the standard B-model camera that had been in use since October 1956. Work on a very-high-resolution camera had begun in early 1963, when the Agency contracted with the Itek Corporation to modify for placement in the U-2 a camera that had been developed for the satellite program. Known as the Delta-II, or the 112A, this device could photograph a 28-kilometer swath with 26° convergent stereoscopic lenses, resulting in a 70° lateral coverage and a ground resolution of 10 inches. This camera was installed in a Detachment H U-2 and used on two missions conducted over Indochina in late December 1963. Resolution was not as high as had been expected, and the unit was returned to Itek for modifications. By early 1964, the 112A had been reworked and was now known as the 112B. In tests it had proved capable of providing photography with resolution in the 10- to 12-inch range.

Detachment G conducted Operation FISH HAWK in May 1964 by sending two pilots, an NPIC photointerpreter, and a U-2G equipped with the Itek 112B camera to make the first operational U-2 flights from an aircraft carrier. On 19 May the U-2 took off from the USS Ranger and overflew the French atomic test area. As soon as the aircraft returned to the Ranger, the film was developed in the carrier's photo lab, and the NPIC photointerpreter then read out the film to see if the photography met the requirements for resolution and quality. A second U-2 flight carried out a similar mission on 22 May. The photography provided all the detail needed to identify the preparations for the nuclear test that occurred later that year.[70]

[70] "U-2 Aircraft Carrier Operations: Project 'WHALE TALE,' Operation FISH HAWK," DS&T, 1964, pp. 17-25.

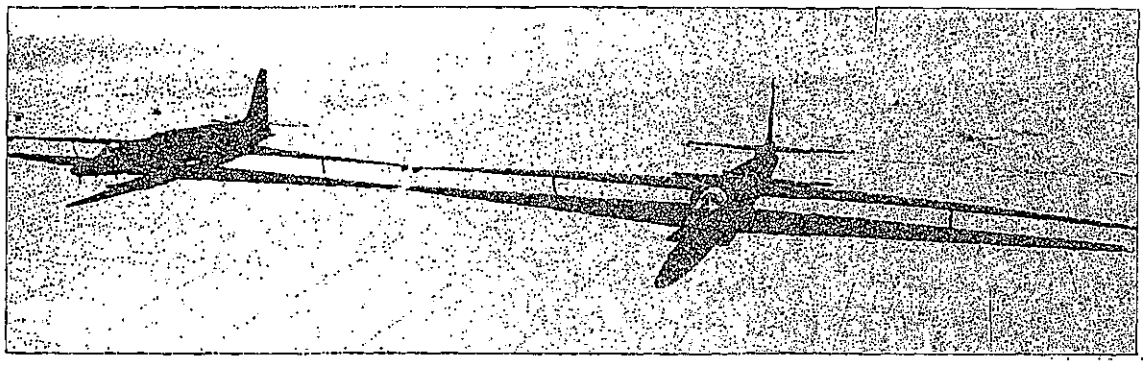
There was never another Agency U-2 mission from an aircraft carrier. Although the idea of using a floating airbase to avoid political sensitivity proved feasible, the cost did not. Aircraft carriers are enormously expensive to operate and require an entire flotilla of vessels to protect and service them. The movement of large numbers of big ships is difficult to conceal and cannot be hastily accomplished, while the deployment of a solitary U-2 to a remote airfield can take place overnight.

A New Version of the U-2

By the summer of 1966, the number of flyable Agency U-2s had dwindled to six—two at Detachment H in Taiwan and four at Detachment G in California—with three more at Lockheed undergoing repair. The Agency had originally ordered 20 U-2s in 1954-55 (the Air Force had purchased another 31 of these planes), and Kelly Johnson's crew at the Skunk Works had managed to assemble four additional craft for the Agency from leftover spare parts and usable sections of crashed aircraft. This brought the total number of U-2s acquired by the Agency to 24, for an average cost of $812,500 each.

At this point, the DCI and the Secretary of Defense on 1 August 1966 decided to place an order with Lockheed for eight more aircraft to be used in the Agency and Air Force U-2 programs—a completely new version of the aircraft. Kelly Johnson had been working on ways to improve the performance of the U-2 since early 1965 because he was concerned that all the modifications and additions to the aircraft over the years had made it so heavy that it had lost almost half of its range and several thousand feet in cruising altitude.[71] The new model, known as the U-2R, had a longer fuselage and a wider wingspan than the original U-2. The U-2R's wings were 103 feet long with 1,000 square feet of lifting surface, in contrast to the U-2C's 80-foot wings with only 600 square feet. The longer fuselage of the U-2R made it possible to provide two pressurized bays with an additional 2.2 cubic meters of equipment space and also achieve a better weight distribution. The net result of all these improvements was a much better performing aircraft. No longer did the U-2 pilot have to worry about keeping the aircraft's speed at altitude within a 6-knot window in the stall/buffet corner of the flight envelope. The envelope was now extended to 20 knots, which greatly improved flyability.

[71] Johnson, "Log for Project X," 2 February 1965, June to October 1965, 20 October 1965; Johnson, "U-2R Log," January to August 1966.



U-2C and U-2R

The U-2R used the upgraded Pratt & Whitney J75/P-13B engine and was able to fly higher—in excess of 74,000 feet—and faster—Mach 0.72 (410 knots), which is 12 knots faster than the U-2C. When flying at the higher altitude, however, the U-2R's range was less than the U-2C's. The restart capability of the P-13B engine was significantly better than the P-13A power plant. As a result, the U-2R could be restarted at 54,000 feet, which was 10,000 feet higher than the U-2C. Francis Gary Powers was one of the Lockheed test pilots who checked out this new aircraft when it first took to the air on 28 August 1967. The last of the U-2Rs was delivered on 11 December 1968.

The increased performance of the U-2R did not come cheaply. At $7.1 million per aircraft, the new models cost almost 10 times as much as the original U-2s. Much of the increased cost was due to inflation, but some was the result of technological advances. The initial order for eight of the new version of the U-2 was followed on 23 November 1966 by an order from the DCI and the Secretary of Defense for four more. This brought the total number of U-2Rs purchased by the CIA and the Air Force to 12.[72]

In addition to a new aircraft, the U-2 program received a new camera. Agency managers felt that, because the B camera was now 10 years old, the U-2R needed a camera that incorporated the many important advances that had occurred in recent years. The 112B—the modified version of the satellite program's stereo camera that had been used in the U-2G—had not proved totally successful. Despite its stereo capability, this camera's shorter focal length could not provide

[72] *OSA History*, chap. 5, pp. 34-36 (TS Codeword); "OSA History-2," chap. 5, pp. 1-2 (TS Codeword).

the scale of imagery needed to obtain the highly technical data desired by analysts. As a result, the Office of Special Activities asked the Hycon Manufacturing Company of Pasadena, California, to adapt its successful high-resolution 48-inch 9- by 9-inch format camera developed for the OXCART aircraft for use in the U-2R. This camera was actually a very advanced version of the original B camera with a new lens designed by James Baker. The new camera was designed to resolve objects smaller than 4 inches.

Hycon began work on the HR-333 camera in 1966. Unlike the OXCART camera, the new unit was to use the split 18- by 18-inch format of the B camera, so the lens had to be redesigned. James Baker's contribution to this effort was a 48-inch f/5.6 system that provided remarkably sharp imagery. Hycon completed the camera in time for it to be installed in the first U-2Rs delivered to the Agency in 1968; it is known as the H camera.[73]

Replacement of the Original U-2s With U-2Rs

As the new U-2Rs began coming off the production line at Lockheed in the autumn of 1968, CIA and the Department of Defense had to decide who would get the new aircraft. At a meeting on 13 November, DCI Richard Helms and Secretary of Defense Robert McNamara agreed that the Air Force and the Agency would each get six U-2Rs. The six older U-2s remaining from the original 1954-55 production were to be kept in flyable condition and be used as replacements if newer models were lost.

Despite the greatly increased capabilities of the new model of the U-2, the era of overflights of hostile territory was over. The U-2R would have six years of useful service with the Agency, but its missions did not include penetration flights over hostile territory.

THE FINAL YEARS OF THE U-2

When the OXCART's brief operational career with the Agency ended in 1968, the U-2 was once again the center of the Agency's manned reconnaissance program. But by this time, reconnaissance aircraft had declined in importance as collection systems. Overflights were a thing of the past. Although Project TACKLE U-2s with Nationalist Chinese

[73] "OSA History-2," chap. 5, pp. 10-12.

pilots were still flying missions targeted against the People's Republic of China, these missions did not overfly PRC territory. Increasingly, Agency U-2s flew missions that did not involve intelligence collection requirements.

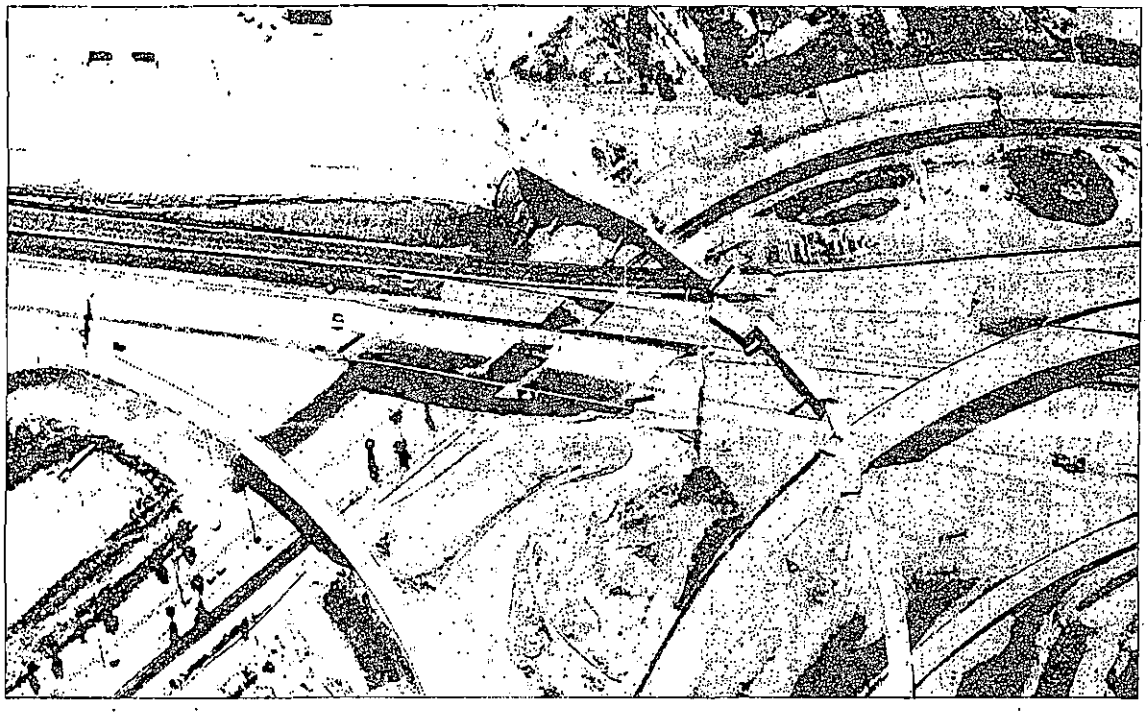
Support to Other Agencies

Beginning in 1964, the Agency conducted a program known as RED DOT for the Department of Defense. RED DOT involved the development and testing of various color, black and white, and infrared films, emulsions, and processing techniques for use in manned and unmanned high-altitude reconnaissance systems. From 1968 until 1974, Detachment G U-2s photographed areas within the United States that were analagous to portions of the Soviet Union in order to test films and techniques for spotting certain targets. This analogous filming was particularly valuable in connection with agricultural areas and nuclear test sites.

Some U-2 missions supported agencies outside the intelligence community. In 1968 and 1969, Detachment G U-2s flew high-altitude photographic missions in conjunction with the Apollo VII and IX spaceflights in response to a NASA request. These flights provided photography of the western United States for comparison with the photography taken by the Apollo crews. The Department of the Interior also requested U-2 support in early 1969 to help determine the extent of damage caused by a leak in an offshore oil well in California's Santa Barbara Channel. After preliminary assessment of the film at NPIC, the mission photography was given to the US Geological Survey for further study.

Also in early 1969, Detachment G began providing coverage of the western United States at the request of the Department of Commerce. U-2s filmed the Sierra snowfield to aid hydrologists in forecasting snowmelt and flooding potentials. Later that year, Detachment G supported the Office of Emergency Preparedness by photographing 61,000 square miles of the southern United States as part of a Hurricane Baseline Survey. These photographs could be used for future damage assessment following a major hurricane. A subsequent mission in fiscal year 1971 continued the Hurricane Baseline Survey by photographing the Gulf Coast. When a major earthquake struck the Los Angeles area on 9 February 1971, Detachment G U-2s flew four sorties to obtain damage-assessment photos.[74]

[74] Ibid., chap. 3, pp. 3-29 (TS Codeword).



Earthquake damage, San Fernando Valley, 1971

Subsequent missions in support of Federal agencies included COMPASS TRIP in fiscal year 1973, when Agency U-2s photographed poppy fields that had been planted by the Bureau of Narcotics in order to provide a standard for comparison with satellite imagery. In the following year, U-2s assisted the Corps of Engineers in conducting a geological survey.

Overseas Deployment Exercises and Missions

With the exception of the Chinese Nationalist-piloted U-2s of Detachment H, all of the Agency's U-2 assets were concentrated in Detachment G in California. To test the ability of Detachment G to respond to a crisis in Europe or the Middle East, the Agency staged an overseas deployment exercise known as SCOPE SAINT each year (unless there was an actual operational deployment, as was the case in 1970, 1973, and 1974). The first of these exercises, SCOPE SAINT–I, took place on 9 October 1968, when Detachment G deployed a U-2G [redacted] The U-2 conducted several training flights and then returned to California. SCOPE SAINT–II followed in April 1969 and demonstrated the feasibility of employing a C-141 aircraft to accompany a

25X1 and 6, E.O.13526

25X1 and 6, E.O.13526

U-2 in flight to its destination. The C-141 carried support equipment to the [redacted] In October 1969, the third of these exercises took place at a different location, [redacted] This exercise demonstrated the continuing British participation in the U-2 program because British pilots flew both the deployment and redeployment legs of the mission.[75]

No overseas deployment exercise was necessary in 1970, for elements of Detachment G actually deployed overseas to provide photography of the Middle East. At the time, President Nixon's National Security Adviser, Henry A. Kissinger, was mediating between the Arabs and Israelis in order to obtain a cease-fire along the Suez Canal, where a virtual undeclared war was taking place. Once agreement was reached in August, Kissinger promised both sides that the United States would monitor the agreed upon 32-mile pullback from the waterway. Originally, Kissinger intended for photosatellites to do the monitoring. One satellite was tasked to photograph the Suez Canal area on 10 August, but the quality of its imagery lacked the detail needed to discover such small targets as gun emplacements and jeeps.

In early August, Kissinger asked the Air Force to provide U-2s to overfly the Canal, but the Air Force demurred, saying it would take several weeks to move a U-2 detachment from Del Rio, Texas, to the Middle East. At this point, DCI Helms told an NSC meeting that the Agency's Detachment G at Edwards Air Force Base could deploy aircraft to [redacted] and begin filming the Suez area within the week, and it did. In fact, the first U-2 arrived in [redacted] only 71 hours after receiving notification to deploy. Between 9 August and 10 November 1970, Agency U-2s flew 29 missions over the cease-fire zone as part of Project EVEN STEVEN. Most flights used the B camera, but 12 were equipped with the new, high-resolution H camera. The EVEN STEVEN U-2s also employed a dozen electronic-intelligence-collection packages, from System-X to System-XXIV. After 10 November 1970, Air Force SR-71s took over the task of photographing the cease-fire zone.[76]

The Middle East was again the cause of a Detachment G deployment in October 1973, when another Arab-Israeli war broke out. Two U-2s deployed to the [redacted] on 7 and 8 October 1973, to be ready for possible coverage of the conflict. Detachment G received no such tasking, however, and the last of the aircraft returned to California on 13 November. The 1973 war did

[75] Ibid., pp. 5-6, 10-11 (~~TS~~ Codeword).

[76] Ibid., pp. 15-20 (~~TS~~ Codeword).

25X1 and 6, E.O.13526

lead to the overseas deployment of Detachment G U-2s in 1974, when the CIA was tasked to monitor the Israeli-Egyptian and later the Israeli-Syrian disengagement areas. On 21 April 1974, a Detachment G U-2 with appropriate support elements arrived at Akrotiri, Cyprus, to conduct Operation OLIVE HARVEST. Between 12 May and 28 July, the detachment conducted six overflights of the disengagement areas. During these missions the electronic warning systems of the U-2 registered numerous radar lockons, but no surface-to-air missiles were fired. On 1 August 1974, responsibility for the OLIVE HARVEST missions as well as the aircraft itself came into the hands of the Air Force as part of the transfer of the entire Agency U-2 program at that time.[77]

The Phaseout of the Office of Special Activities

The Agency's U-2 program had been under review since the autumn of 1969 to determine if it should be continued along with the larger Air Force U-2 program. In December 1969, President Nixon decided to keep the Agency's program in existence through 1971 and asked for a formal review by the 40 Committee (the new name for the 303 Committee/Special Group). In August 1970, the committee recommended continuing the program through fiscal year 1972. On 12 August 1972, the 40 Committee again favored continuation of the CIA U-2 program. This recommendation was motivated primarily by a desire not to alienate the Nationalist Chinese Government by eliminating Project TACKLE. In June 1973, however, DCI James R. Schlesinger informed the 40 Committee that this project could be terminated without causing major difficulties with the Nationalist Chinese. On 30 August 1973, the 40 Committee approved the CIA's plans to terminate the U-2 program effective 1 August 1974. The Air Force would assume funding responsibility for the four U-2R aircraft assigned to the Agency and would take physical possession of them then or shortly thereafter. On 1 April 1974, Ambassador Walter P. McConaughy informed the Nationalist Chinese Government of the US intention to end the U-2 project, and the two countries then worked out a schedule for phasing out Project TACKLE.[78]

The transfer of all Agency U-2s to the Air Force eliminated Detachments G and H. Their parent organization, the Office of Special Activities, began its phaseout immediately thereafter. The 20-year career of the U-2 with the CIA had come to an end.

---

[77] Ibid., pp. 31-34 (~~CRS~~ Codeword).

[78] Ibid., chap. 10, pp. 1-4 (~~CRS~~ Codeword).

The U-2's Intended Successor: Project OXCART, 1956-1968

Before the U-2 became operational in June 1956, CIA project officials had estimated that its life expectancy for flying safely over the Soviet Union would be between 18 months and two years. After overflights began and the Soviets demonstrated the capability of tracking and attempting to intercept the U-2, this estimate seemed too optimistic. By August 1956, Richard Bissell was so concerned about the U-2's vulnerability that he despaired of its ability to avoid destruction for six months, let alone two years.

To extend the U-2's useful operational life, project officials first attempted to reduce the aircraft's vulnerability to detection by Soviet radars. Project RAINBOW's efforts to mask the radar image of the U-2 not only proved ineffective, but actually made the aircraft more vulnerable by adding extra weight that reduced its maximum altitude. Because Soviet radar operators continued to find and track U-2s equipped with antiradar systems, the CIA canceled Project RAINBOW in May 1958.

Long before the failure of Project RAINBOW, Richard Bissell and his Air Force assistant, Col. Jack A. Gibbs, had begun to look for a more radical solution to the problem of Soviet radar detection—an entirely new aircraft. In the late summer of 1956, the two officials visited a number of airframe contractors in a search for new ideas. Among the more unusual was Northrop Aviation's proposal for a gigantic aircraft with a very-high-lift wing. Because it would not be made of metal, the wing would require a type of bridge truss on its upper side to give it rigidity. The proposed aircraft would achieve

altitudes of 80,000 to 90,000 feet but only at subsonic speeds, just enough to keep it airborne.[1]

The slow-flying Northrop design did not solve the problem of radar detection, and in 1957 the emphasis switched to supersonic designs. In August 1957, the Scientific Engineering Institute (SEI), a CIA proprietary firm that had been working on ways to reduce the U-2's vulnerability to radar, began to investigate the possibility of designing an aircraft with a very small radar cross section. SEI soon discovered that supersonic speed greatly reduced the chances of detection by radar.[2] From this point on, the CIA's attention focused increasingly on the possibility of building an aircraft that could fly at both extremely high speeds and high altitudes while incorporating the best ideas in radar-absorbing or radar-deflecting techniques.

THE EVALUATION OF DESIGNS FOR A SUCCESSOR TO THE U-2

By the autumn of 1957, Bissell and Gibbs had collected so many ideas for a successor to the U-2 that Bissell asked DCI Dulles for permission to establish an advisory committee to assist in the selection process. Bissell also felt that the support of a committee of prominent scientists and engineers would prove useful when it came time to ask for funding for such an expensive project. Edwin Land became the chairman of the new committee, which included some of the scientists and engineers who had served on previous advisory bodies for overhead reconnaissance: Edward Purcell, Allen F. Donovan, H. Guyford Stever, and Eugene P. Kiefer. The Air Force's chief scientist, Courtland D. Perkins, was also a member. The committee first met in November 1957 and held six more meetings between July 1958 and the late summer of 1959. The meetings usually took place in Land's Boston office and almost always included the Air Force's Assistant Secretary for Research and Development, Dr. Joseph V. Charyk, and his Navy counterpart, Garrison Norton. Designers from several aircraft manufacturers also attended some of the meetings.[3]

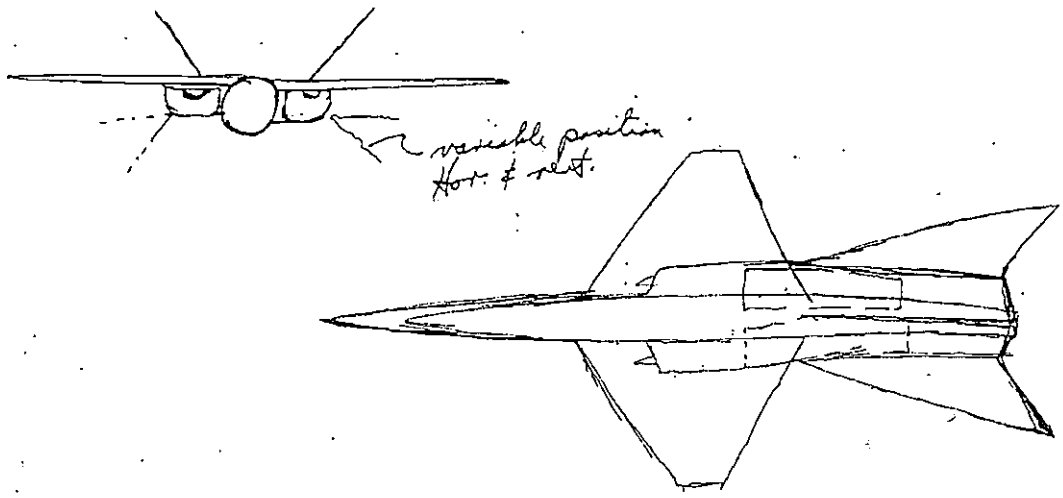
---

[1] Donovan interview (S).

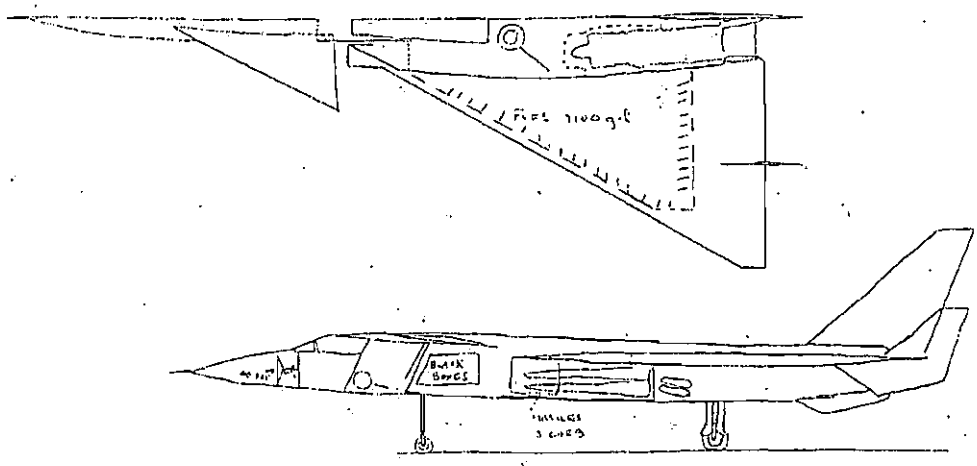
[2] Thomas P. Minnich, "The OXCART Story," *Studies in Intelligence* 15 (Winter 1971):2 (S).

[3] Clarence L. Johnson, Report No. SP-1362, "History of the OXCART Program," Lockheed Aircraft Corporation, Burbank, CA, 1 July 1968, p. 1 (TS Codeword).

A-1, 23 April 1958



A-1, 26 June 1958



Johnson's first drawing of the "U-3" (A-1); revised version of the A-1

The two most prominent firms involved in the search for a new aircraft were Lockheed, which had designed the successful U-2, and Convair, which was building the supersonic B-58 "Hustler" bomber for the Air Force and also working on an even faster model known as the B-58B "Super Hustler." Early in 1958, Richard Bissell asked officials from both firms to submit designs for a high-speed reconnaissance aircraft. During the spring and summer of 1958, both firms worked on design concepts without government contracts or funds.

Following extended discussions with Bissell on the subject of a supersonic successor to the U-2, Lockheed's Kelly Johnson began designing an aircraft that would cruise at Mach 3.0 at altitudes above 90,000 feet. On 23 July 1958, Johnson presented his new high-speed concept to Land's advisory committee, which expressed interest in the approach he was taking. At the same meeting, Navy representatives presented a concept for a high-altitude reconnaissance vehicle that examined the possibility of developing a ramjet-powered, inflatable, rubber vehicle that would be lifted to altitude by a balloon and then be propelled by a rocket to a speed where the ramjets could produce thrust. Richard Bissell asked Johnson to evaluate this concept, and three weeks later, after receiving more details from Navy representatives, Kelly Johnson made some quick calculations that showed that the design was impractical because the balloon would have to be a mile in diameter to lift the vehicle, which in turn would need a wing surface area greater than one-seventh of an acre to carry the payload.[4]

By September 1958, Lockheed had studied a number of possible configurations, some based on ramjet engines, others with both ramjets and turbojets. Personnel at Lockheed's Skunk Works referred to these aircraft concepts as "Archangel-1," "Archangel-2," and so forth, a carryover from the original nickname of "Angel" given to the U-2 during its development. These nicknames for the various designs soon became simply "A-1," "A-2," etc.

In September 1958, the Land committee met again to review all the concepts then under consideration and to winnow out the few that were most practicable. Among the concepts rejected were the Navy's proposal for an inflatable, ramjet-powered aircraft, a Boeing proposal for a 190-foot-long hydrogen-powered inflatable aircraft, and a

[4] Clarence L. Johnson, "Development of the Lockheed SR-71 Blackbird," *Studies in Intelligence* 26 (Summer 1982):4 (U); Johnson, "Archangel log," 23 July 1958, 14 August 1958.

Lockheed design for a hydrogen-powered aircraft (the CL-400). The committee examined two other Kelly Johnson designs at this meeting—a tailless subsonic aircraft with a very-low-radar cross section (the G2A) and a new supersonic design (the A-2)—and did not accept either one, the former because of its slow speed and the latter because of its dependence on exotic fuels for its ramjets and its overall high cost. The committee approved the continuation of Convair's work on a ramjet-powered Mach 4.0 "parasite" aircraft that would be launched from a specially configured version of the B-58B bomber. The design was termed a parasite because it could not take off on its own but needed a larger aircraft to carry it aloft and accelerate it to the speed required to start the ramjet engine. The Convair design was called the FISH.[5]

Two months later, after reviewing the Convair proposal and yet another Lockheed design for a high-speed reconnaissance aircraft (the A-3), the Land committee concluded in late November 1958 that it would indeed be feasible to build an aircraft whose speed and altitude would make radar tracking difficult or impossible. The committee, therefore, recommended that DCI Dulles ask President Eisenhower to approve further pursuit of the project and to provide funds for additional studies and tests.[6]

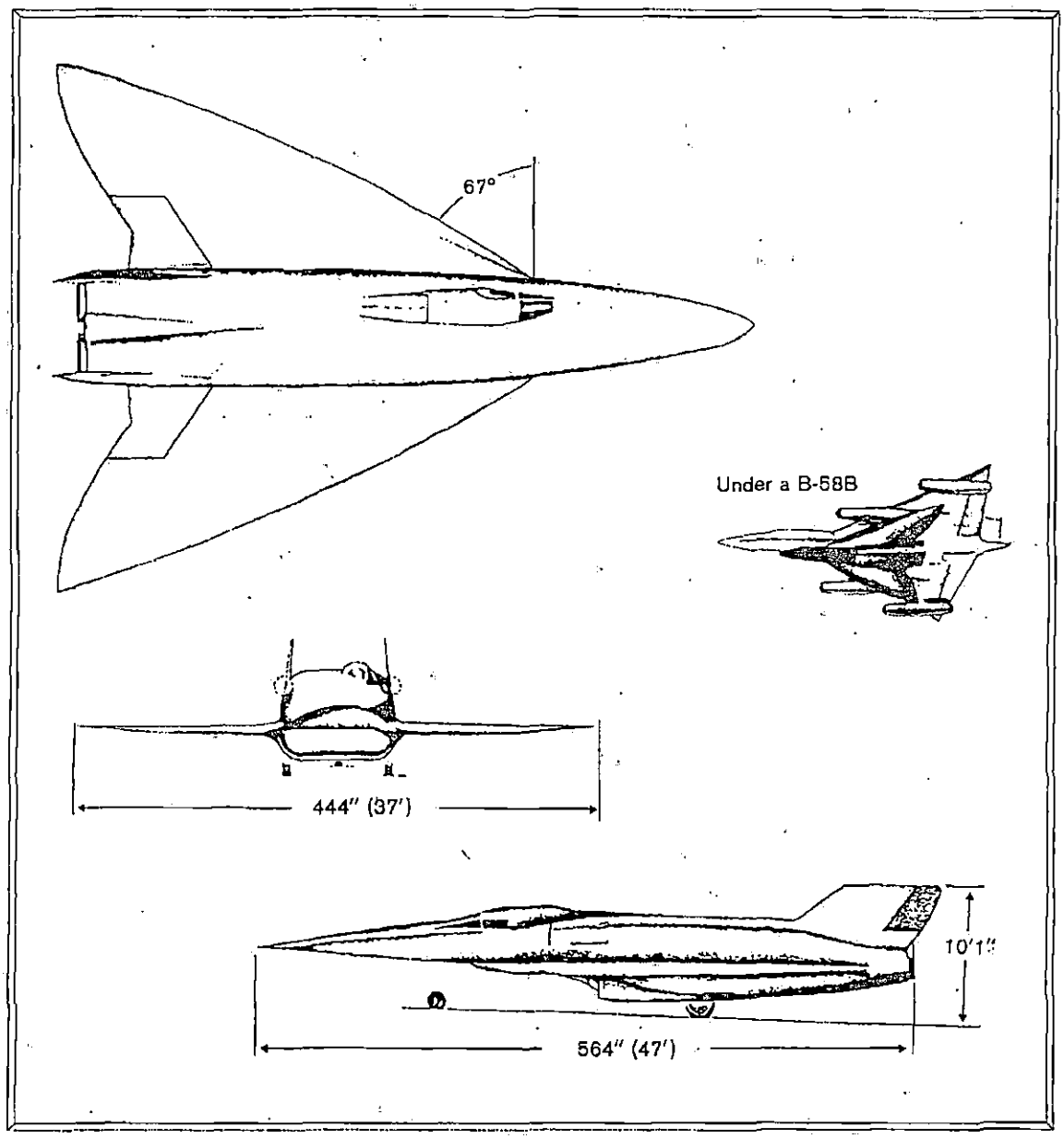
On 17 December 1958, Allen Dulles and Richard Bissell briefed the President on the progress toward a successor to the U-2. Also present were Land and Purcell from the advisory committee, Presidential Science Adviser James Killian, and Air Force Secretary Donald Quarles. DCI Dulles reviewed the results of the U-2 missions to date and stated his belief that a successor to the U-2 could be used all over the world and "would have a much greater invulnerability to detection."

Bissell then described the two competing projects by Lockheed and Convair, noting that the chief question at the moment was whether to use air launch or ground takeoff. The next phase, he added, would be detailed engineering, at the end of which it was proposed that 12 aircraft be ordered at a cost of about $100 million.

---

[5] *OSA History*, chap. 20, p. 8 (~~CPS~~ Codeword); Johnson, "Archangel log," 17-24 September 1958.

[6] *OSA Chronology*, p. 21 (~~CPS~~ Codeword); Minnich, "OXCART Story," p. 3 (S); *OSA History*, chap. 20, p. 8 (~~CPS~~ Codeword); Johnson, "Archangel log," 12 November 1958.



Convair FISH

Kelly Johnson's A-2 Design

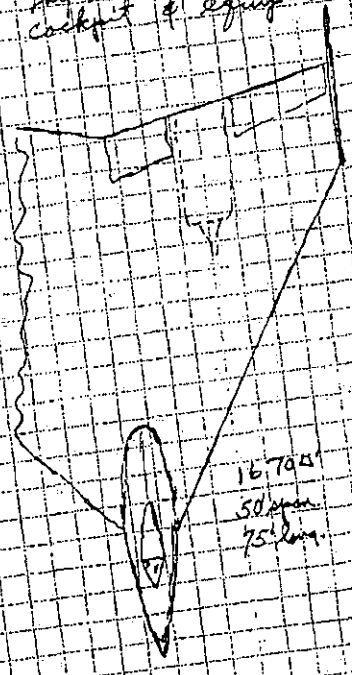
LOCKHEED AIRCRAFT CORP.
CALIFORNIA DIVISION

Prepared: Sept 20, 58 CW
Checked:
Approved:
Date:
Title: Try a Brown Job.
Page: 11
Model:
Report No.:

Use 10,000# airplane at 135,000 —

Try to eliminate fuselage except for cockpit & equip bay

L/D = 6.0 to 5.5



Wing @ 1.32#	= 2200#
Tail (incl elev)	+ 600
Engines - 2 @ 800#	1600
Gear (required now)	400
Instruments	60
Surface controls	800
Power supply	200
Electronics	150
Air Cond.	250
Oxygen	50
Subtotal =	5810
Cockpit & Bay	= 1000
Ejectseat & gun	= 150
Payload	= 500
Subtotal	= 7460
Fuel system - tanks	= 500
	7960

Fuel load = 2000#

Range = $575 \times 6.0 \times \frac{3.0}{2.10} \times \frac{10,000}{8000}$ = 1090 miles

Kelly Johnson's A-3 Design

Sept. 29 '58
CW

LOCKHEED AIRCRAFT CORP.

Design of A-3.

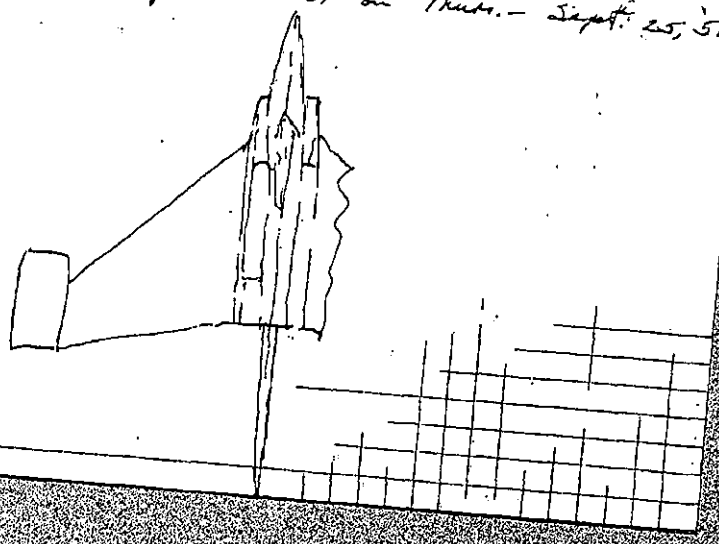
1

From previous work — a basic design of the following characteristics was derived (AAL trip - 3 page)

Area — 500 □'
~~Area~~ wt. — 17,000 #
Wt. at 100,000' — 13,200
2 — A.B. JT-12A
2 — 30" Ram jets.
300 # payload.
M = 3.0 @ 100,000' (3.2?)

Basic concept — reduce radar C.S.

Data given to SP on Thurs. — Sept. 25, '58



Although President Eisenhower supported the purchase of this type of aircraft, he questioned the plan to procure any before they had been tested. Promising that more thought would be given to the matter before such an order was placed, Secretary Quarles noted that CIA, the Defense Department, and the Bureau of the Budget were working on a funding plan for the project. The President suggested that the Air Force "could support the project by transferring some reconnaissance money." At the close of the meeting, Eisenhower asked the group to return after completing the next work phase to discuss further stages of the project with him.[7]

COMPETITION BETWEEN LOCKHEED AND CONVAIR

With funding for the proposed new type of aircraft now available, Richard Bissell asked Lockheed and Convair to submit detailed proposals. During the first half of 1959, both Lockheed and Convair worked to reduce the radar cross section of their designs, with assistance from Franklin Rodgers of the Scientific Engineering Institute. In pursuing his antiradar studies, Rodgers had discovered a phenomenon that he believed could be used to advantage by the new reconnaissance aircraft. Known as the Blip/Scan Ratio but also referred to as the Rodgers' Effect, this phenomenon involved three elements: the strength of a radar return, the altitude of the object being illuminated by the radar, and the persistence of the radar return on the radar screen (Pulse-Position Indicator display).

Most tracking radars in the late 1950s swept a band of sky 30° to 45° wide and 360° in circumference. Any object encountered in this area reflected the radar pulse in a manner directly proportional to its size—the larger the object, the stronger the returning radar signal. This return appeared on the cathode-ray tube of the radar screen as a spot or blip, and the persistence of this blip on the radar screen also depended on the strength of the radar return, with blips from larger objects remaining on the screen longer. During the late 1950s and early 1960s, a human radar operator watched the radar screen and kept track of the blips that indicated aircraft within the radar's field of view.

[7] Andrew J. Goodpaster, "Memorandum of Conference with the President, 17 December 1958, 10:26 a.m.," 22 December 1958, WHOSS, Alpha, DDEL ~~(TS)~~.

Rodgers determined that a high-altitude object moving two to three times as fast as a normal aircraft would produce such a small blip with so little persistence that the radar operator would have great difficulty tracking it, if indeed he could even see it. Rodgers estimated that for an aircraft to take advantage of this Blip/Scan Ratio phenomenon it must fly at altitudes approaching 90,000 feet and have a radar cross section of less than 10 square meters, preferably not much over 5 square meters. However, for a Mach 3.0 aircraft to achieve such a small radar cross section, its designers would have to make many concessions in its structural design and aerodynamics.[8]

By the summer of 1959, both firms had completed their proposals. In early June, Lockheed submitted a design for a ground-launched aircraft known as the A-11. It would have a speed of Mach 3.2, a range of 3,200 miles, an altitude of 90,000 feet, and a completion date of January 1961. Kelly Johnson had refused to reduce the aerodynamics of his design in order to achieve a greater antiradar capability, and the A-11's radar cross section, although not great, was substantially larger than that of the much smaller parasite aircraft being designed by Convair.[9]

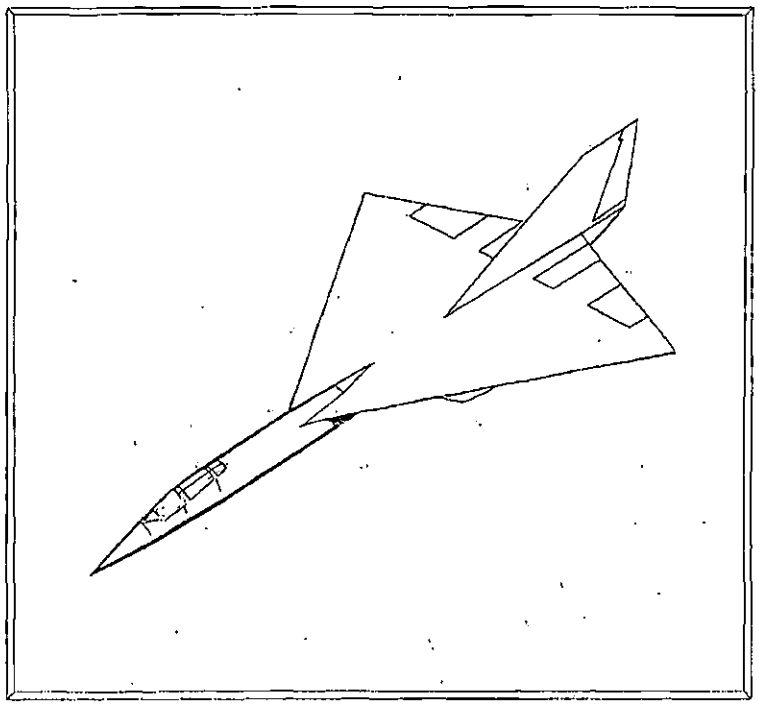
The Convair proposal called for a small, manned, ramjet-powered, reconnaissance vehicle to be air launched from one of two specially configured Convair B-58B Super Hustlers. The FISH vehicle, a radical lifting body with a very-small-radar cross section, would fly at Mach 4.2 at 90,000 feet and have a range of 3,900 miles. Two Marquardt ramjets would power its Mach 4.2 dash over the target area. Once the FISH decelerated, two Pratt & Whitney JT-12 turbojets would bring it back to base. The ramjet exit nozzles and wing edges would be constructed of Pyroceram, a ceramic material that could withstand the high temperatures of very high speeds and would absorb radiofrequency energy from radar pulses. Convair stated that the FISH could be ready by January 1961.[10]

Convair's proposal depended on two uncertain factors. First and foremost was the unproven technology of ramjet engines. At the time, no aircraft in existence could carry a large, ramjet-powered craft into the sky and then accelerate to sufficient speed for the ramjet engines

[8] Unnumbered Convair document on the Blip/Scan Ratio or Rodgers' Effect ~~(TS)~~.

[9] Johnson, "Archangel log," December 1958–July 1959.

[10] *OSA History*, chap. 20, p. 12 (TS Codeword); Convair Division, General Dynamics Corporation, "Project FISH Status Review," 9 June 1959 ~~(S)~~.



Lockheed A-11

to be ignited. Since ramjet engines had only been tested in wind tunnels, there was no available data to prove that these engines would work in the application proposed by Convair. The second uncertain factor was the B-58B bomber that was supposed to achieve Mach 2.2 before launching the FISH above 35,000 feet. This version of the B-58 was still in the design stage.

Convair's proposal suffered a major setback in June 1959, when the Air Force canceled the B-58B project. Conversion of the older, slower B-58A into a supersonic launching platform for the FISH was ruled out by the high cost and technical difficulties involved. Moreover, the Air Force was unwilling to part with two aircraft from the small inventory of its most advanced bomber. Even had the B-58B program not been canceled, however, the FISH proposal would probably not have been feasible. Convair engineers had calculated that the added weight of the FISH would prevent the B-58B from achieving the speed required to ignite the parasite aircraft's ramjet engines.

The Convair proposal was therefore unusable, but the Lockheed design with its high radar cross section was also unacceptable to the Land committee. On 14 July 1959, the committee rejected both

designs and continued the competition. Lockheed continued to work on developing a design that would be less vulnerable to detection, and Convair received a new CIA contract to design an air-breathing twin-engine aircraft that would meet the general specifications being followed by Lockheed.[11]

Following recommendations by the Land committee, both Lockheed and Convair incorporated the Pratt & Whitney J58 power plant into their designs. This engine had originally been developed for the Navy's large, jet-powered flying boat, the Glenn L. Martin Company's P6M Seamaster, and was the most powerful engine available. In 1958 the Navy had canceled the Seamaster program, which had left Pratt & Whitney without a buyer for the powerful J58 engine.[12]

Although the Land committee had not yet found an acceptable design, it informed President Eisenhower on 20 July 1959 that the search was making good progress. Concerned about the U-2's vulnerability to detection and possible interception and aware that the photosatellite project was encountering significant problems, the President gave his final approval to the high-speed reconnaissance aircraft project.[13]

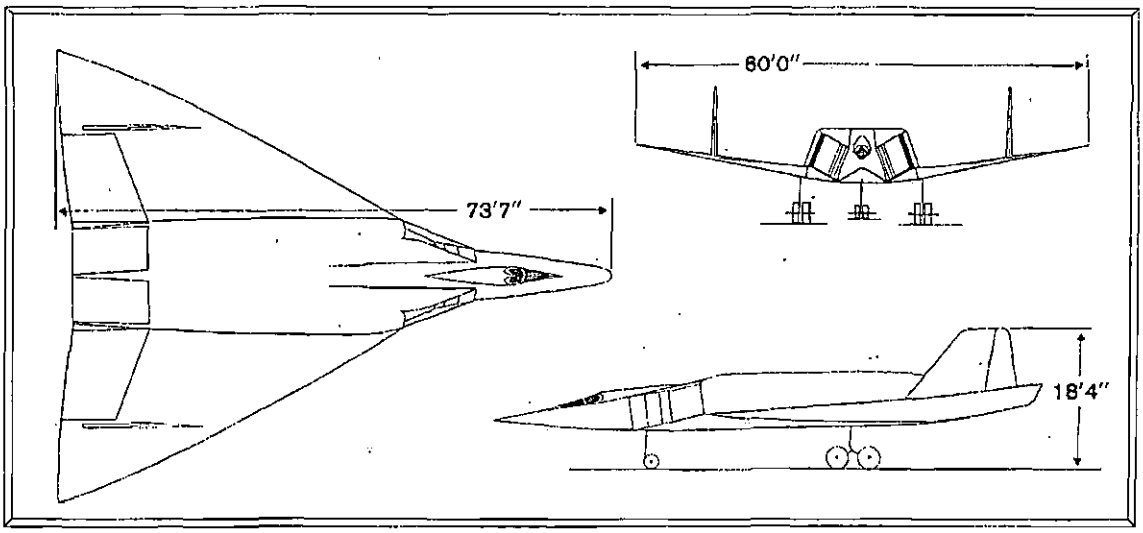
THE SELECTION OF THE LOCKHEED DESIGN

By the late summer of 1959, both Convair and Lockheed had completed new designs for a follow-on to the U-2. Convair's entry, known as the KINGFISH, used much of the technology developed for the F-102, F-106, and B-58, including stainless steel honeycomb skin, planiform wing design, and a crew capsule escape system, which eliminated the need for the pilot to wear a pressurized suit. The KINGFISH had two side-by-side J58 engines inside the fuselage, which significantly reduced the radar cross section. Two additional

[11] *OSA History*, chap. 20, p. 15 (TS Codeword).

[12] Cunningham interview, 4 October 1983 (TS Codeword); Joseph V. Charyk, interview by Donald E. Welzenbach, tape recording, Washington, DC, 5 December 1984 (~~TS~~ Codeword).

[13] Andrew J. Goodpaster, "Memorandum of Conference with the President," 20 July 1959, WHOSS, ALPHA, DDEL (~~TS~~).



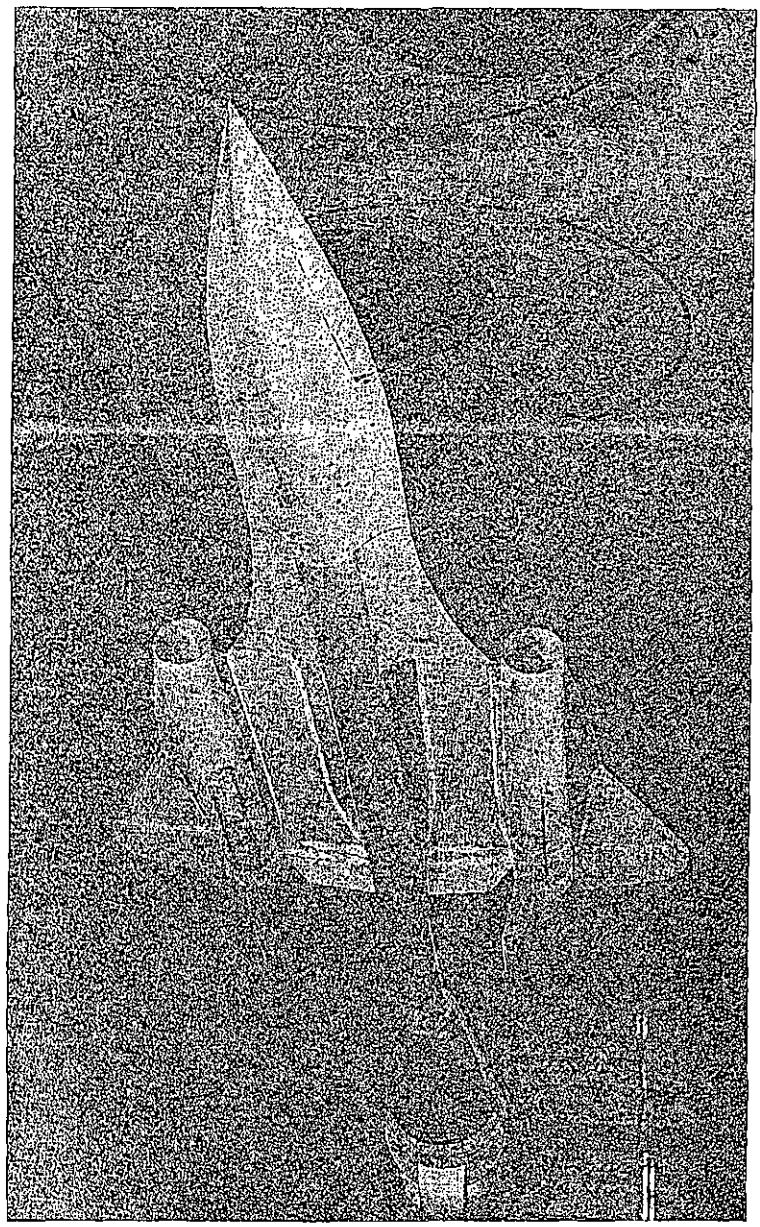
Convair KINGFISH

important design features that contributed to a small radar return were fiberglass engine inlets and wings whose leading edges were made of Pyroceram.[14]

Lockheed's new entry was much like its first, but with several modifications and a new designator, A-12. It, too, would employ two of the powerful J58 engines. Lockheed's major innovation in reducing radar return was a cesium additive in the fuel, which decreased the radar cross section of the afterburner plume. This improvement had been proposed by Edward Purcell of the Land committee. Desiring to save weight, Kelly Johnson had decided not to construct the A-12 out of steel. Traditional lightweight metals such as aluminum were out of the question because they could not stand the heat that would be generated as the A-12 flew at Mach 3.2, so Johnson chose a titanium alloy.

On 20 August 1959, Lockheed and Convair submitted their proposals to a joint Department of Defense, Air Force, and CIA selection panel. As the table shows, the two aircraft were similar in performance

[14] Convair Division, General Dynamics Corporation, "KINGFISH Summary Report," 1959 (S). Kelly Johnson was very skeptical of the Convair design, noting in the Archangel project log on 1-20 August 1959: "Convair have promised substantially reduced radar cross section on an airplane the size of our A-12. They are doing this, in my view, with total disregard for aerodynamics, inlet and afterburner performance."



Wind tunnel test of A-12 model

characteristics, although the Lockheed design's specifications were slightly better in each category. The Lockheed design was also preferable in terms of overall cost. In the vital area of vulnerability to radar detection, however, the Convair design was superior. Its smaller size and internally mounted engines gave it a smaller radar cross section than the Lockheed A-12.[15]

Comparison of Lockheed and Convair Designs

	Lockheed A-12	Convair KINGFISH
Speed	Mach 3.2	Mach 3.2
Range (total)	4,120 nm	3,400 nm
Range (at altitude)	3,800 nm	3,400 nm
Cruising Altitude		
Start	84,500 ft.	85,000 ft.
Midrange	91,000 ft.	88,000 ft.
End	97,600 ft.	94,000 ft.
Cost summary (for 12 aircraft without engines)	$96.6 million	$121.6 million

Some of the CIA representatives initially favored the Convair KINGFISH design because of its smaller radar cross section, but they were eventually convinced to support the Lockheed design by the Air Force members of the panel, who believed that Convair's cost overruns and production delays on the B-58 project might be repeated in this new project. In contrast, Lockheed had produced the U-2 under budget and on time. Another factor favoring the A-12 was security. Lockheed had experience in running a highly secure facility (the Skunk Works) in which all of the key employees were already cleared by the Agency.

Despite its vote in favor of the Lockheed proposal, the selection panel remained concerned about the A-12's vulnerability to radar detection and therefore required Lockheed to prove its concept for reducing the A-12's radar cross section by 1 January 1960. On 14 September 1959, the CIA awarded a four-month contract to Lockheed

[15] *OSA History*, chap. 20, pp. 18-19 (~~[illegible]~~ Codeword).

to proceed with antiradar studies, aerodynamic structural tests, and engineering designs. This research and all later work on the A-12 took place under a new codename, Project OXCART, established at the end of August 1959 to replace its more widely known predecessor, Project GUSTO.[16] The CIA's project manager for OXCART was John Parangosky, who had long been associated with the U-2 program.

EFFORTS TO REDUCE THE A-12'S RADAR CROSS SECTION

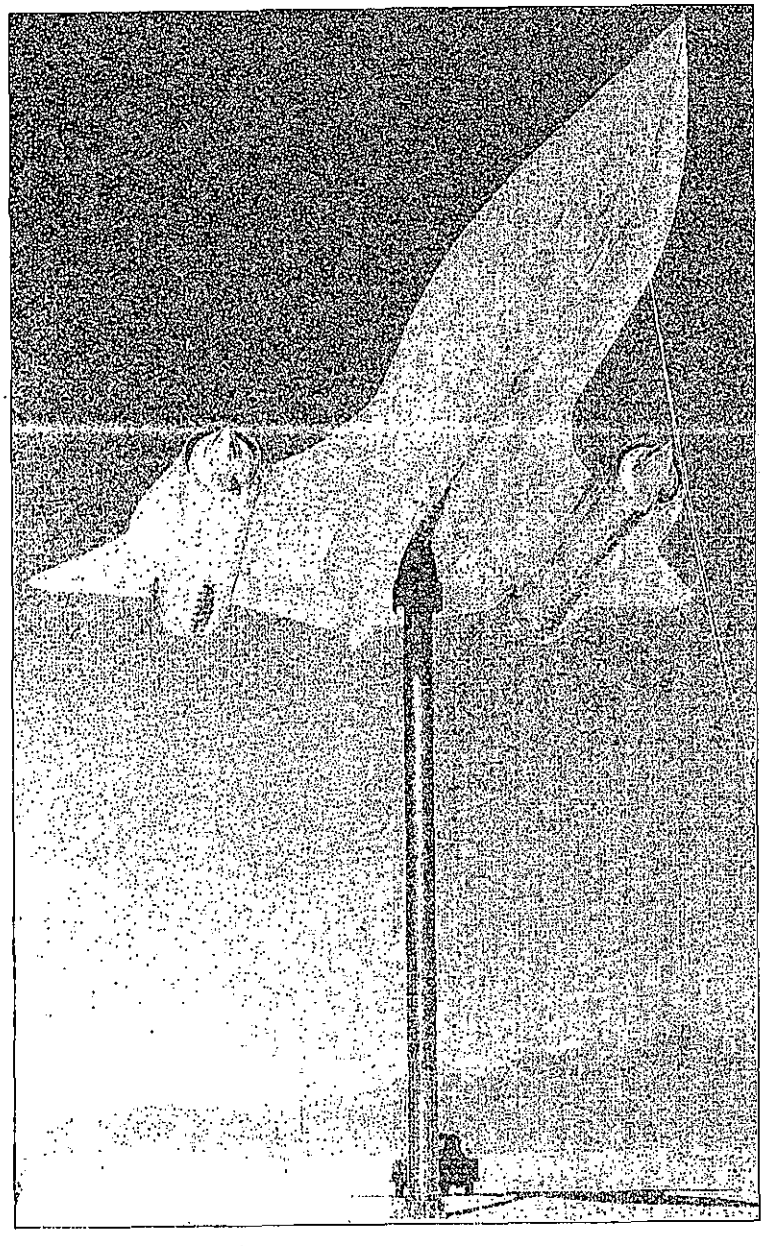
During the spring of 1959, Kelly Johnson's Skunk Works crew—which then numbered only 50—had begun building a full-scale mockup of the proposed aircraft. The mockup was to be tested for its radar cross section by Edgerton, Germeshausen & Grier (EG&G) in cooperation with the Scientific Engineering Institute at a small testing facility at Indian Springs, Nevada. Lockheed objected to this site because its pylon would not support the full-scale mockup and because the facilities were in full view of a nearby highway. On 10 September 1959, EG&G agreed to move its radar test facility to the former U-2 testing site at Area 51 of the Atomic Energy Commission's Nevada Proving Grounds.[17]

When the new radar test facility with its larger pylon was ready, Johnson put the A-12 mockup on a specially designed trailer truck that carried it from Burbank to Area 51. By 18 November 1959, the mockup was in place atop the pylon, and radar testing could begin. These tests soon proved that Lockheed's concept of shape, fuel additive, and nonmetallic parts was workable, but it would take more than 18 months of testing and adjustment before the OXCART achieved a satisfactory radar cross section.

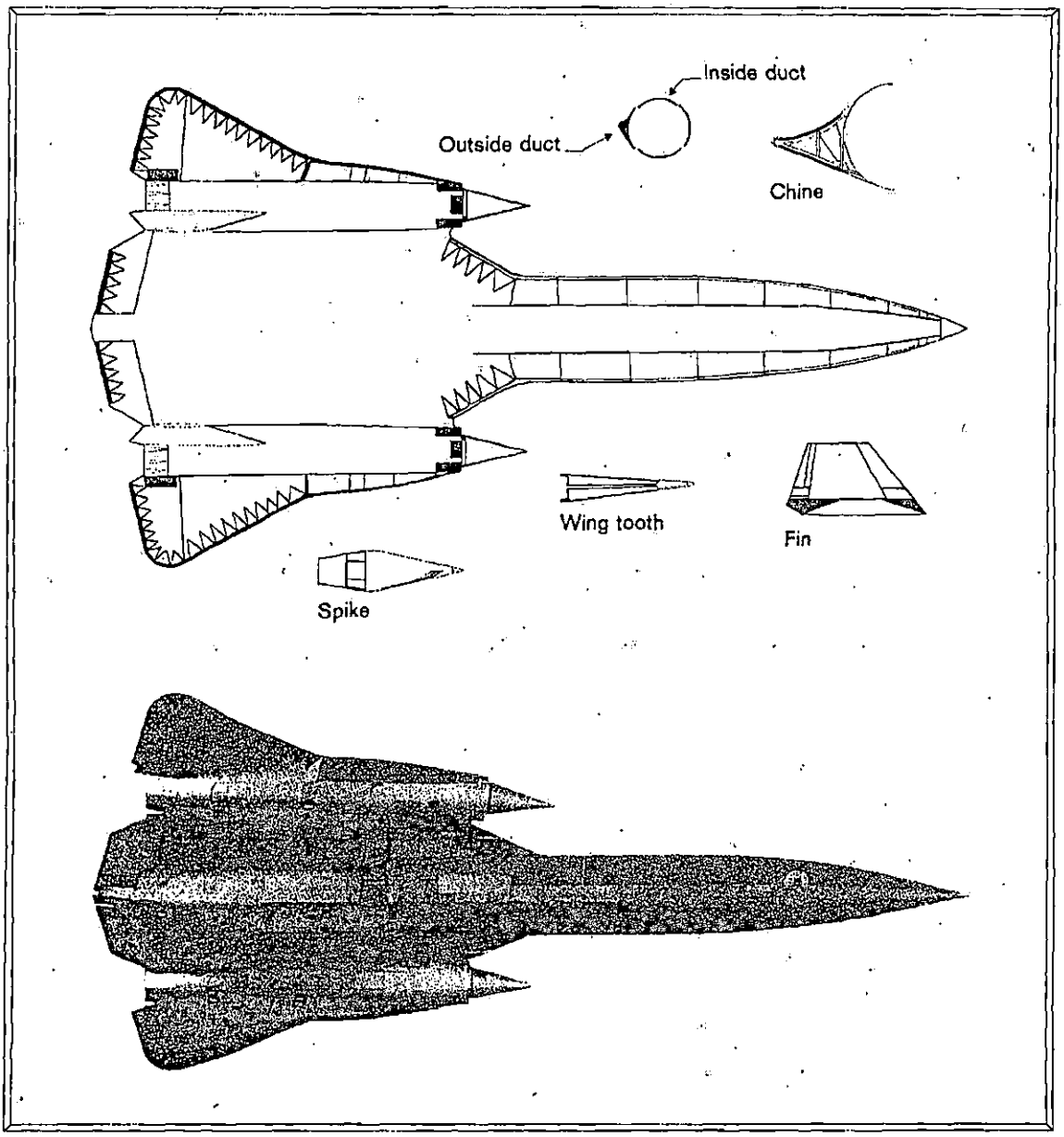
It was in the course of this radar testing that the OXCART received its characteristic cobra-like appearance. Edward Purcell and Franklin Rodgers had come up with a theory that a continuously curving airframe would be difficult to track with a radar pulse because it would present few corner reflectors or sharp angles from which pulses could bounce in the direction of the radar. To achieve the continuously curving airframe, Kelly Johnson added thin, curved extensions to the engine housings and leading edges of the wings and

[16] Parangosky interview (S); *OSA History*, chap. 20, pp. 19-21 (~~TS~~ Codeword).

[17] *OSA History*, chap. 20, p. 22 (~~TS~~ Codeword).



Radar testing of A-12 mockup



Antiradar features of the A-12

eventually to the fuselage itself, creating what is known as a chine on each side. At first Johnson was concerned that these additions might impair the airworthiness of the plane, but wind tunnel testing determined that the chines actually imparted a useful aerodynamic lift to the vehicle. Because titanium was very brittle and therefore difficult to bend, Johnson achieved the necessary curvature by combining triangular-shaped pieces of titanium called fillets. These fillets were glued to the framework of the chines with a special adhesive, epoxy resin.

On later OXCART models the fillets were made from electrically resistive honeycomb plastic with a glass-fiber surface that would not melt at high speed. When struck by a radar pulse, the composite chines tended to absorb the pulse rather than reflect it. A similar approach was used for the leading edges of the wings. Again electrically resistive honeycomb material was fabricated into triangular shapes, known as wing teeth, and fitted into the titanium wings. Both the metal and composite fillets and teeth were held in place with the newly developed epoxy cements.

The greatest remaining area of concern in the A-12's radar cross section was the two vertical stabilizers. To reduce radar reflections, Kelly Johnson canted the stabilizers inward 15° and fabricated them out of resin-impregnated nonmetallic materials. Once these changes were completed, the only metal in each vertical stabilizer was a stainless steel pivot. The Air Force, which later ordered several versions of the OXCART aircraft for its own use, never adopted the laminated vertical stabilizers.[18]

THE OXCART CONTRACT

By mid-January 1960, Lockheed had demonstrated that its concept of shape, fuel additive, and nonmetallic parts would reduce the OXCART's radar cross section substantially. Richard Bissell, however, was very upset to learn that the changes had led to a reduction in the aircraft's performance, which meant it would not be able to attain the penetration altitude he had promised to President Eisenhower. Kelly Johnson then proposed to reduce the aircraft's weight by 1,000 pounds and increase the fuel load by 2,000 pounds, making it possible

[18] Johnson, "Development of Lockheed SR-71," pp. 6-7; *OSA History*, chap. 20, p. 35 (~~TK~~ Codeword).

to achieve the desired target altitude of 91,000 feet. Afterward, he noted in the project log: "We have no performance margins left; so this project, instead of being 10 times as hard as anything we have done, is 12 times as hard. This matches the design number and is obviously right." [19]

These changes satisfied Bissell, who notified Johnson on 26 January that the CIA was authorizing the construction of 12 of the new aircraft. The actual contract was signed on 11 February 1960. Lockheed's original quotation for the project was $96.6 million for 12 aircraft, but technological difficulties eventually made this price impossible to meet. Recognizing that fabricating an aircraft from titanium might involve unforeseen difficulties, the CIA included a clause in the contract that allowed costs to be reevaluated. During the next five years, this clause had to be invoked on a number of occasions as the A-12's costs soared to more than double the original estimate.[20]

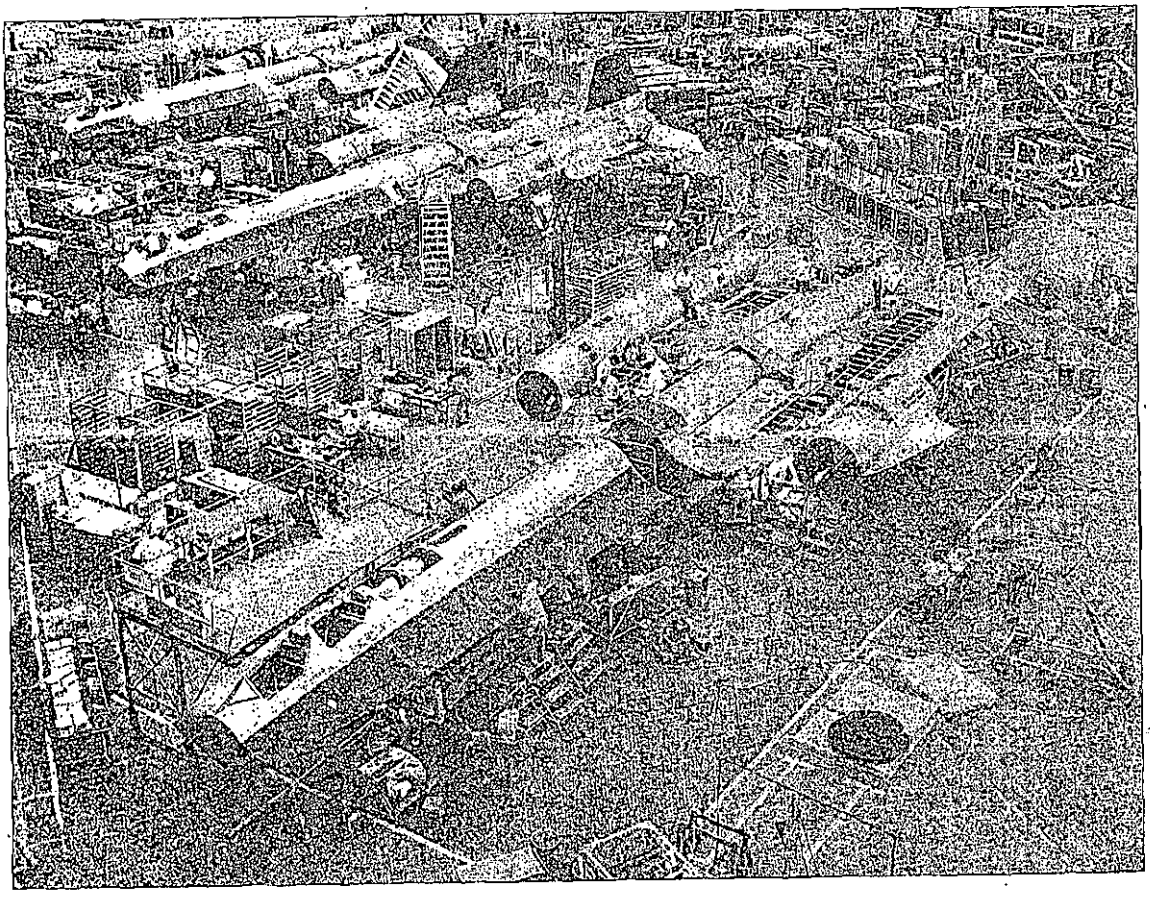
NEW TECHNOLOGIES NECESSITATED BY OXCART'S HIGH SPEED

According to the specifications, the OXCART aircraft was to achieve a speed of Mach 3.2 (2,064 knots or 0.57 miles per second, which would make it as fast as a rifle bullet), have a range of 4,120 nautical miles, and reach altitudes of 84,500 to 97,600 feet. The new aircraft would thus be more than five times as fast as the U-2 and would go almost 3 miles higher.

One major disadvantage of the OXCART's great speed was high temperatures. Flying through the earth's atmosphere at Mach 3.2 heated portions of the aircraft's skin to almost 900°F. An aircraft operating at these high speeds and high temperatures required fuels, lubricants, and hydraulic fluids that had not yet been invented. The OXCART's fuel requirement called for a low-vapor-pressure fuel with a low volume at operating temperatures; the fuel would also be used as a heat sink to cool various parts of the aircraft. The J58 engines required lubricants that did not break down at the very high operating temperatures of Mach 3.2 speeds. This requirement led to the

[19] Johnson, "Archangel log," 21 January 1960.

[20] *OSA History*, chap. 20, pp. 27-29, 33-34, 36 (~~TS~~ Codeword).

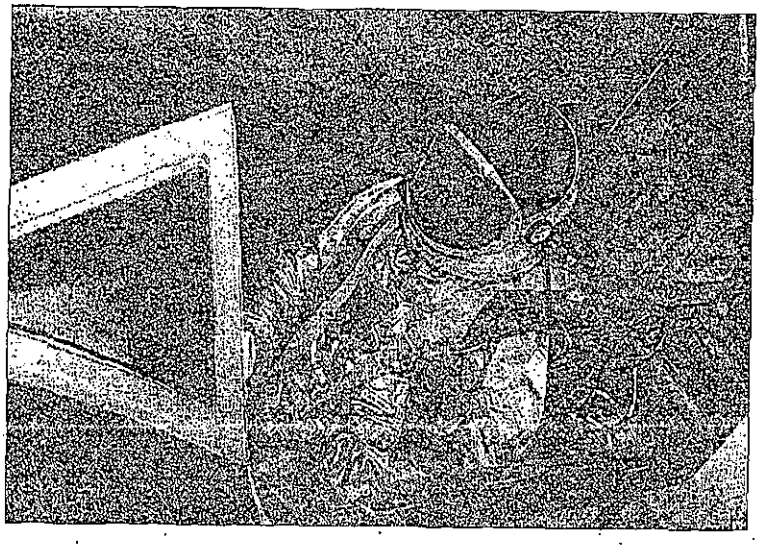


OXCART production facilities

invention of synthetic lubricants. Lockheed also had to search long and hard for a hydraulic fluid that would not vaporize at high speed but would still be usable at low altitudes. Finding a suitable hydraulic pump was just as difficult. Kelly Johnson finally modified a pump that was being developed for North American's B-70 bomber project.[21]

Some of the greatest problems related to the high speeds and high temperatures at which the OXCART operated resulted from working with the material chosen for the airframe—titanium. After evaluating many materials, Johnson had chosen an alloy of titanium

[21] Johnson, "Development of Lockheed SR-71," pp. 11-12.



OXCART pilot suit

(B-120) characterized by great strength, relatively light weight, and good resistance to high temperatures, but high in cost. As strong as stainless steel, titanium weighed slightly more than half as much. Obtaining sufficient quantities of titanium of a quality suitable for fabricating aircraft components proved very difficult because methods for maintaining good quality control during the milling of titanium were not fully developed. Up to 80 percent of the early deliveries from Titanium Metals Corporation had to be rejected. It was not until 1961, when company officials were informed of the objectives and high priority of the OXCART program, that problems with the titanium supply ended. Even after sufficient high-quality titanium was received, Lockheed's difficulties with the metal were not over. Titanium was so hard that tools normally used in aircraft fabrication broke; new ones therefore had to be devised. Assembly line production was not possible, and the cost of the program mounted well above original estimates.[22]

The high temperatures that the OXCART would encounter also necessitated planning for the pilot's safety and comfort because the inside of the aircraft would be like a moderately hot oven. To save

[22] Minnich, "OXCART Story," p. 5 (S); *OSA History*, chap. 20, p. 33 (TS Codeword).

weight, Kelly Johnson did not attempt to insulate the interior of the aircraft. The pilot would therefore have to wear a type of space suit with its own cooling, pressure control, oxygen supply, and other necessities for survival.

DESIGNING THE OXCART'S CAMERAS

Providing cameras for the A-12 posed a number of unique problems. In late 1959, OXCART managers asked Perkin-Elmer, Eastman Kodak, and Hycon to develop three different photographic systems for the new aircraft. These cameras would provide a range of photography from high-ground-resolution stereo to extremely-high-resolution spotting data.

The Perkin-Elmer (P-E) entry, known as the Type-I camera, was a high-ground-resolution general stereo camera using an f/4.0 18-inch lens and 6.6-inch film. It produced pairs of photographs covering a swath 71 miles wide with an approximately 30-percent stereo overlap. The system had a 5,000-foot film supply and was able to resolve 140 lines per millimeter and provide a ground resolution of 12 inches.

To meet severe design constraints in the areas of size, weight, thermal environment, desired photographic resolution, and coverage, Perkin Elmer's Dr. Roderick M. Scott employed concepts never before used in camera systems. These included the use of a reflecting cube rather than a prism for the scanner, a concentric film supply and takeup system to minimize weight shift, a constant-velocity film transport that provided for the contiguous placement of stereo images on one piece of film, and airbars for the film transport and takeup systems.[23]

Eastman Kodak's entry, called the Type-II camera, was a high-convergent stereo device using a 21-inch lens and 8-inch film. It produced pairs of photographs covering a swath 60 miles wide with an approximately 30-percent stereo overlap. It had an 8,400-foot film supply and was able to resolve 105 lines per millimeter and provide a ground resolution of 17 inches.

[23] *OSA History*, chap. 20, p. 26 (~~TK~~ Codeword); Miannich, "OXCART Story," p. 4 (S).

The Hycon entry, designed by James Baker and known as the Type-IV camera, was a spotting camera with extremely-high-ground resolution. In fact, it was an advanced version of the highly reliable B camera developed for the original U-2 program. It used a 48-inch Baker-designed f/5.6 lens to focus images onto 9.5-inch film. Like the B camera it could provide seven frames of photography covering a swath 41 miles wide with stereo overlap on 19 miles of the swath. The Hycon camera carried the largest film supply of the three cameras, 12,000 feet. It was able to resolve 100 lines per millimeter and provide a ground resolution of 8 inches. A version of this 48-inch Hycon camera, known as the H camera, later saw service in U-2R aircraft.

Each of the three camera systems had unique capabilities and advantages, so all three were purchased for the OXCART. Before they could be effectively employed in the aircraft, however, new types of camera windows were needed. The OXCART's camera windows had to be completely free from optical distortion. Achieving this goal was difficult in a window whose exterior would be subjected to temperatures of 550°F while the interior surface would be only 150°F. After three years and the expenditure of $2 million in research and development, the Corning Glass Works, which had joined this effort as a Perkin-Elmer subcontractor, solved the problem of producing a camera window that could withstand tremendous heat differentials. Its quartz glass window was fused to the metal frame by an unprecedented process involving high-frequency sound waves.[24]

Later in the program, the OXCART received yet another camera system. In 1964 the Texas Instruments Corporation developed an infrared camera for Project TACKLE U-2s that were being used to determine whether the People's Republic of China was producing weapons-grade nuclear material. This stereo device, known as the FFD-4, was adapted for use in OXCART. The camera had an effective focal length of 50 inches and a 150-foot supply of 3.5-inch film. The camera's resolution was 3°C thermally, 1 milliradian spatially, and 60 feet on the ground. It could be used for both day and night imagery collection.

[24] Baker interview (S); Minnich, "OXCART Story," pp. 5-6 (S).

CHOOSING PILOTS FOR OXCART

Just as in the U-2 program, the Air Force provided considerable support to Project OXCART, including training, fuel storage, and weather service. One of the most important areas of support was the provision of pilots; all of the OXCART pilots came from the Air Force. Prospective pilots had to be qualified in the most advanced fighters and be emotionally stable and well motivated. In contrast to 1955, when cover considerations had limited the U-2 pilot selection process to individuals with reserve commissions, the Air Force was able to devise personnel and cover procedures that enabled both regular and reserve officers to volunteer to become OXCART pilots. Because of the limited size of the A-12 cockpit, they had to be under six feet tall and weigh less than 175 pounds. Following extensive physical and psychological screening, 16 potential nominees were selected for intensive security and medical screening by the Agency. By the end of this screening in November 1961, only five individuals had been approved and had accepted the Agency's offer of employment on a highly classified project involving a very advanced aircraft. A second search and screening raised the number of pilots for the OXCART to eleven. The thorough screening process produced an elite group of pilots; all but one of these 11 officers eventually became generals. The new pilots transferred from military to civilian status and received compensation and insurance arrangements somewhat better than those of the U-2 pilots.[25]

SELECTION OF A TESTING SITE FOR THE OXCART

From the very beginning, it was clear that Lockheed could not test the OXCART aircraft at its Burbank facility, where the runway was too short and too exposed to the public. The ideal testing site would be far removed from metropolitan areas, away from civil and military airways, easily accessible by air, blessed with good weather, capable of accommodating large numbers of personnel, near an Air Force installation, and having a runway at least 8,000 feet long. But no such place was to be found.

After considering 10 Air Force bases programmed for closing, Richard Bissell decided to upgrade the Area 51 site in Nevada where the U-2 had been tested. Although its personnel accommodations, fuel

[25] Minnich, "OXCART Story," pp. 6-7 (~~S~~); *OSA History*, chap. 20, pp. 48-50 (~~TS~~ Codeword); Geary interview with Pedlow (~~S~~).

storage capacity, and runway length were insufficient for the OXCART program, the site's remote location would greatly ease the task of maintaining the program's security, and a moderate construction program could provide adequate facilities. Construction began in September 1960; a C-47 shuttle service ferried work crews from Burbank to Las Vegas and from Las Vegas to the site.

The new 8,500-foot runway was completed by 15 November 1960. Kelly Johnson had been reluctant to have a standard Air Force runway with expansion joints every 25 feet because he feared the joints would set up undesirable vibrations in the speedy aircraft. At his suggestion a 150-foot wide runway was therefore constructed of six 25-foot-wide longitudinal sections, each 150 feet long but staggered. This layout put most of the expansion joints parallel to the direction of aircraft roll and reduced the frequency of the joints.

Additional improvements included the resurfacing of 18 miles of highway leading to the base so that heavy fuel trucks could bring in the necessary fuel. The need for additional buildings on the base was met by the Navy. Three surplus Navy hangars were dismantled, moved, and reassembled on the north side of the base, and more than 100 surplus Navy housing buildings were also transported to Area 51. All essential facilities were ready in time for the forecast delivery date of the first A-12 on 1 August 1961.[26]

Unfortunately, this delivery date began to slip further and further into the future. Delays in obtaining the titanium, and later the J58 engines, caused the postponement of the final assembly of the first plane. Eventually, Kelly Johnson and Agency project officials decided to begin testing without waiting for the J58 engines by using Pratt & Whitney J75/19W engines, designed for the Convair F-106, to test the A-12 at altitudes up to 50,000 feet and at speeds up to Mach 1.6. Such a change, however, meant that the engine compartment of the first aircraft had to be reconfigured to accommodate the J75 engine. Lockheed hoped that this substitution would permit the delivery of the first A-12 by 22 December 1961 and its initial test flight by 27 February 1962.

Lockheed ran into so many technological problems with the OXCART effort that by October 1961 its costs had swollen to $136 million and were still climbing. Something obviously had to be done

[26] *OSA History*, chap. 20, pp. 39-40, 43, 51 (~~TS~~ Codeword); Minnich, "OXCART Story," pp. 7-9 (~~S~~).

to reduce expenditures. After much refiguring, project officials decided to decrease the number of deliverable aircraft. Amendment No. 11 to the contract reduced from 12 to 10 the number of A-12s, for a total cost of $161.2 million.[27]

The cancellation of these two A-12s was offset by an Air Force order for the development of a supersonic interceptor variant of the A-12 to serve as a replacement for the North American F-108A Rapier interceptor project, which had been canceled in late 1960. With the assistance of the Agency's west coast contracting office, the Air Force entered into an agreement with Lockheed to produce three AF-12 aircraft, based on the A-12 design but modified to carry a second crewman and three air-to-air missiles. This effort was called Project KEDLOCK. The AF-12 (later redesignated the YF-12A) was designed to intercept enemy bombers long before they reached the United States, and initial Air Force plans envisioned a force of up to 100 of these supersonic interceptors. In fact, only three of these planes were built and delivered during the 1963-64 time frame because Secretary of Defense McNamara canceled the program as a cost-cutting measure. The Air Force bore all of the costs of the YF-12A project; CIA was only involved in helping to write "black" contracts.[28]

Lockheed was not the only OXCART contractor having trouble containing costs; Pratt & Whitney was fighting an even bigger battle. In mid-1961, Pratt & Whitney overruns threatened to halt the entire OXCART project. At the suggestion of Cdr. William Holcomb in the office of the Chief of Naval Materiel, Richard Bissell asked the Navy to assist in funding the J58's development. After hearing Bissell and Holcomb's suggestion that the J58 might be used in future Navy aircraft, VAdm. William A. Schoech, Chief of the Navy Materiel Command that had originally financed the J58 engine, authorized the transfer of $38 million in end-of-year funds to the project, thus keeping the OXCART's head above water.[29] As it turned out, the J58 was never used in a Navy aircraft.

---

[27] *OSA History*, chap. 20, pp. 46-47, 51-55 (~~TS~~ Codeword); Minnich, "OXCART Story," p. 10 (S).

[28] *OSA History*, chap. 20, pp. 46-47 (~~TS~~ Codeword).

[29] Parangosky interview (S); *OSA History*, chap. 20, p. 55 (~~TS~~ Codeword). During this period, Kelly Johnson was very disappointed with Pratt & Whitney's work on the J58, particularly when they shocked him in September 1961 with the news that the engine would be overweight, underpowered, and late. Johnson, "Archangel log," 11 September 1961.

DELIVERY OF THE FIRST OXCART

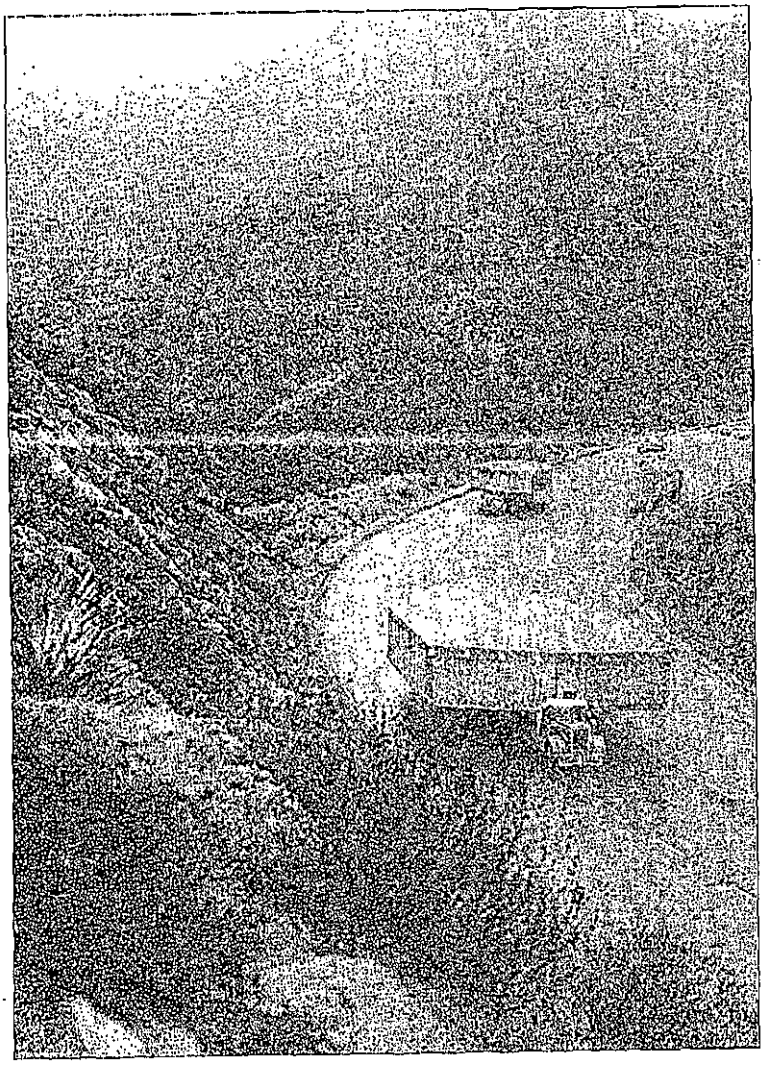
The first A-12, known as article 121, was assembled and tested at Burbank during January and February 1962. Since it could not be flown to the Nevada site, the aircraft had to be partially disassembled and put on a specially designed trailer that cost nearly $100,000. The entire fuselage, without the wings, was crated and covered, creating a load 35 feet wide and 105 feet long. To transport this huge load safely over the hundreds of miles to the site, obstructing road signs were removed, trees were trimmed, and some roadbanks had to be leveled. The plane left Burbank on 26 February 1962 and arrived at Area 51 two days later.

After the fuselage arrived in Nevada, its wings were attached and the J75 engines were installed, but the aircraft was still not ready to be tested. This new delay was caused by leaking fuel tanks, a problem that would never be solved completely. Because the A-12's high speeds heat the titanium airframe to more than 500°F, Lockheed designers had to make allowances for expansion. When the metal was cold, the expansion joints were at their widest. In the fuel tanks, these gaps were filled by pliable sealants, but the fuel for the A-12's engines acted as a strong reducing agent that softened the sealants, causing leaks. Thus, when fuel was first poured into the aircraft, 68 leaks developed. Lockheed technicians then stripped and replaced all the sealant, a tedious and time consuming procedure because the sealant required four curing cycles, each at a different temperature over a period of 30 to 54 hours. The engineers were never able to discover a sealant compound that was completely impervious to the jet fuel while remaining elastic enough to expand and contract sufficiently. The A-12's tanks continued to leak, so when it was fueled, it only received enough fuel to get airborne. The plane would then rendezvous with a tanker, top off its tanks, and immediately climb to operating altitude, causing the metal to expand and the leaks to stop.[30]

CHANGES IN THE PROJECT MANAGEMENT

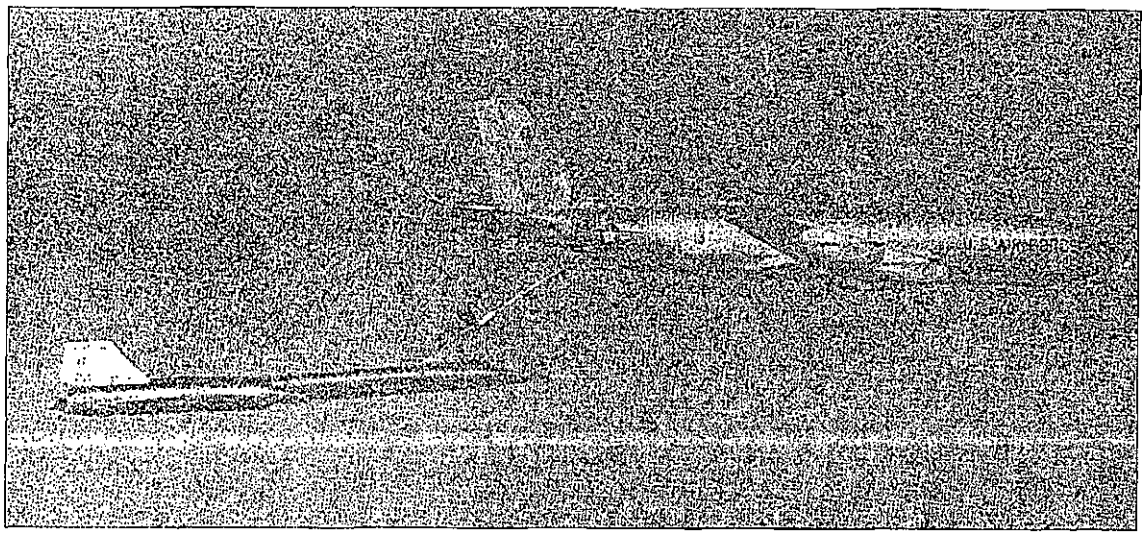
Richard Bissell, whose concern for the viability of the U-2 in 1956 had led to the establishment of Project OXCART and who had directed its growth all along, was no longer in charge when the first

[30] *OSA History*, chap. 20, p. 62 (~~J5~~ Codeword); Minnich, "OXCART Story," p. 11 (~~S~~).



Delivery of OXCART aircraft to Area 51

OXCART aircraft took to the air. He resigned from the Agency in February 1962, and his departure brought a major reorganization of the reconnaissance program. The Development Projects Division of the Directorate of Plans, with its two aircraft (OXCART and U-2) and its satellite project, were transferred to the new Directorate of Research headed by Herbert (Pete) Scoville. The following year Scoville resigned and this Directorate was reorganized and its name changed to the Directorate of Science and Technology, with Albert (Bud) Wheelon, Jr. as its first head. The overhead reconnaissance



In-flight refueling of the OXCART

projects belonged to the Office of Special Activities, headed by Col. Jack C. Ledford, who now had the title of Assistant Director for Special Activities. These project management changes in the CIA had no immediate impact on the OXCART project because the aircraft was still in the development stage, handled mainly by the contractors. Moreover, a good deal of continuity was provided by officers who had served for a number of years with the U-2 program and were now involved with OXCART: James Cunningham, the Deputy Assistant Director for Special Activities; Col. Leo Geary, the Air Force's project officer for the two aircraft; and John Parangosky, who oversaw the day-to-day affairs of the OXCART project.

OXCART'S FIRST FLIGHTS

With new sealant in its fuel tanks, the prototype OXCART was ready to take to the air. On 25 April 1962, test pilot Louis Schalk took "article 121" for an unofficial, unannounced flight, which was an old Lockheed tradition. He flew the craft less than two miles at an altitude of about 20 feet and encountered considerable problems because of the improper hookup of several controls. These were promptly repaired and on the next day, 26 April, Schalk made the official 40-minute maiden flight. After a beautiful takeoff, the aircraft began shedding the triangular fillets that covered the framework of the chines along the edge of the aircraft body. The lost

fillets, which had been secured to the airframe with epoxy resin, had to be recovered and reaffixed to the aircraft, a process that took the next four days.

Once the fillets were in place, the OXCART's official first flight took place on 30 April 1962, witnessed by a number of Agency personnel including DDR Scoville. Richard Bissell was also present, and Kelly Johnson noted in the project log, "I was very happy to have Dick see this flight, with all that he has contributed to the program."[31] This official first flight was also the first flight with the wheels up. Piloted again by Schalk, the OXCART took off at 170 knots and climbed to 30,000 feet. During the 59-minute flight, the A-12 achieved a top speed of 340 knots. Kelly Johnson declared it to be the smoothest first test flight of any aircraft he had designed or tested. On 2 May 1962, during the second test flight, the OXCART broke the sound barrier, achieving a speed of Mach 1.1.[32]

John Parangosky

Four more aircraft, including a two-seat trainer, arrived at the testing site before the end of the year. During the second delivery on 26 June 1962, the extra-wide vehicle carrying the aircraft accidentally struck a Greyhound bus traveling in the opposite direction. Project managers quickly authorized payment of $4,890 for the damage done to the bus in order to avoid having to explain in court why the OXCART delivery vehicle was so wide.

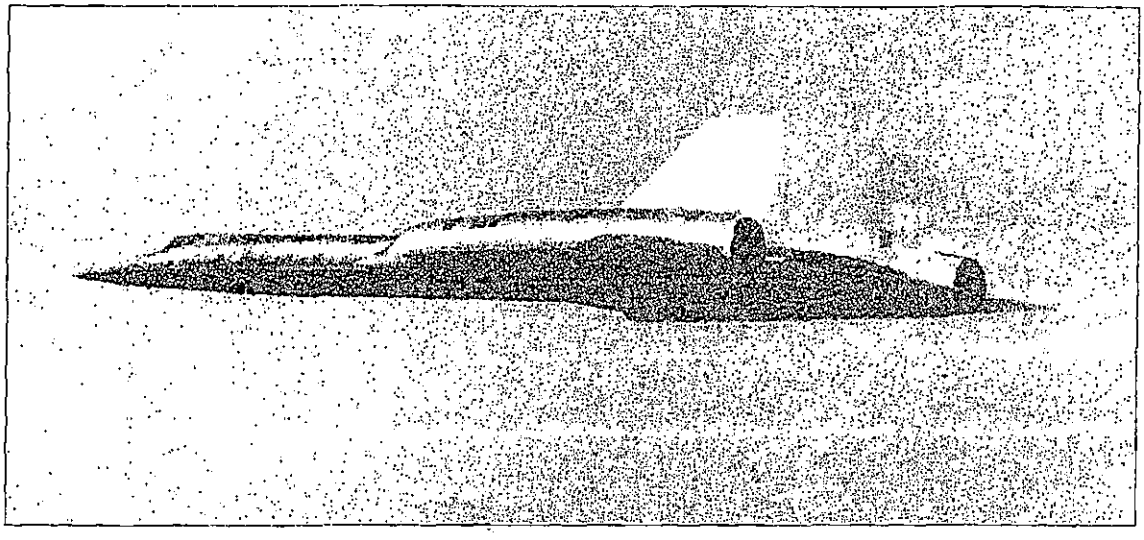
One of the biggest problems connected with flight testing the A-12 was keeping its existence secret. Realizing that the nation's air traffic controllers would be among the first unwitting people to learn about the plane, the Deputy Assistant Director for Special Activities, James Cunningham, had called on Federal Aviation Administrator Najeeb E. Halaby in early 1962 to brief him about the craft's existence and ask his assistance in keeping it secret. Halaby cooperated fully with the Agency and personally briefed all FAA regional chiefs on how to handle reports of unusually fast, high-flying aircraft. Air controllers were warned not to mention the craft on the radio but to submit written reports of sightings or radar trackings. The Air Force gave similar briefings to NORAD, the North American Air Defense Command.[33]

---

[31] Johnson, "Archangel log," 30 April 1962.

[32] *OSA History*, chap. 20, p. 63 (~~CPS~~ Codeword); Minnich, "OXCART Story," pp. 11-12 (S).

[33] Minnich, "OXCART Story," pp. 10-11 (S); *OSA History*, chap. 20, p. 60 (~~CPS~~ Codeword).



First flight of the A-12, 30 April 1962

Initial testing could not explore the A-12's maximum potential, since the J58 engine was still not ready. Developing this power plant to OXCART specifications was proving much more difficult than had been expected because the J58 had to reach performance levels never before achieved by a jet engine, while operating under extremely difficult environmental conditions. To simulate the stress that the J58 would undergo during maximum power output (Mach 3.2 at 97,000 feet), the power plant was tested in the exhaust stream of a J75 engine. In the course of this extremely severe testing, the J58's problems were gradually overcome. By January 1963, Pratt & Whitney had delivered 10 J58 engines to the Nevada testing site. The first flight of an A-12 with two J58 engines took place on 15 January 1963.[34]

SPEED-RELATED PROBLEMS

As J58-equipped A-12s reached higher and higher speeds, more difficulties arose. Major problems developed at speeds between Mach 2.4 and 2.8 because the aircraft's shock wave interfered with the flow of air into the engine, greatly reducing its performance. Solving this problem required long and often highly frustrating experimentation

[34] William H. Brown, "J58/SR-71 Propulsion Integration," *Studies in Intelligence* 26 (Summer 1982):pp. 17-18 (U); *OSA History*, chap. 20, pp. 58, 67 (~~TS~~ Codeword).

that ultimately required a complete redesign of the air-inlet system that controlled the amount of air admitted to the engine. In the new, adjustable inlet the cone-shaped projection at the front—known as a spike—was designed to move in or out as much as three feet in order to capture and contain the shock wave produced by the aircraft at high speeds, thus preventing the shock wave from blowing out the fire inside the engine.[35]

Another J58 engine problem in early 1963 was foreign object damage. Small objects such as pens, pencils, screws, bolts, nuts, and metal shavings that fell into the engine nacelles during assembly at Burbank were sucked into the power plant during initial engine testing at Area 51 and damaged impeller and compressor vanes. To control the problem Lockheed instituted a program that included X-rays, shaking of the nacelles, installing screens over various air inlets to the engine, and even having workers wear coveralls without breast pockets. Another source of foreign object damage was trash on the runways. The giant J58 engines acted like immense vacuum cleaners, sucking in anything lying loose on the paving as they propelled the A-12 down the runway for takeoff. To prevent engine damage, Area 51 personnel had to sweep and vacuum the runway before aircraft takeoff.[36]

NEW VERSIONS OF THE OXCART

In 1962 the Agency and the Air Force ordered two more versions of the OXCART (in addition to the A-12 and the YF-12A). One was a modification of the A-12 to carry and launch ramjet-powered, 43-feet-long drones capable of reaching Mach 3.3 The two-seater mothership received the designation M-12; the drone was called the D-21. This project was known as TAGBOARD. The original development of the drones and mothership was sponsored by the CIA, but in June 1963 the project was turned over to the Air Force, which had overall responsibility for unmanned reconnaissance aircraft. Development of the M-12/D-21 combination continued until 1966, when an unsuccessful D-21 launch caused the loss of its mothership and the death of one of the crew members. Afterward the Air Force turned to B-52 bombers to carry the drones.[37]

---

[35] *OSA History*, chap. 20, p. 67 (~~GRS~~ Codeword).

[36] Johnson, "Development of Lockheed SR-71," p. 12.

[37] *OSA History*, chap. 20, p. 71; Jay Miller, *Lockheed SR-71 (A12/YF12/D-21)*, Aerofax Minigraph 1 (Arlington, Texas: Aerofax, Inc., 1985), p. 3.

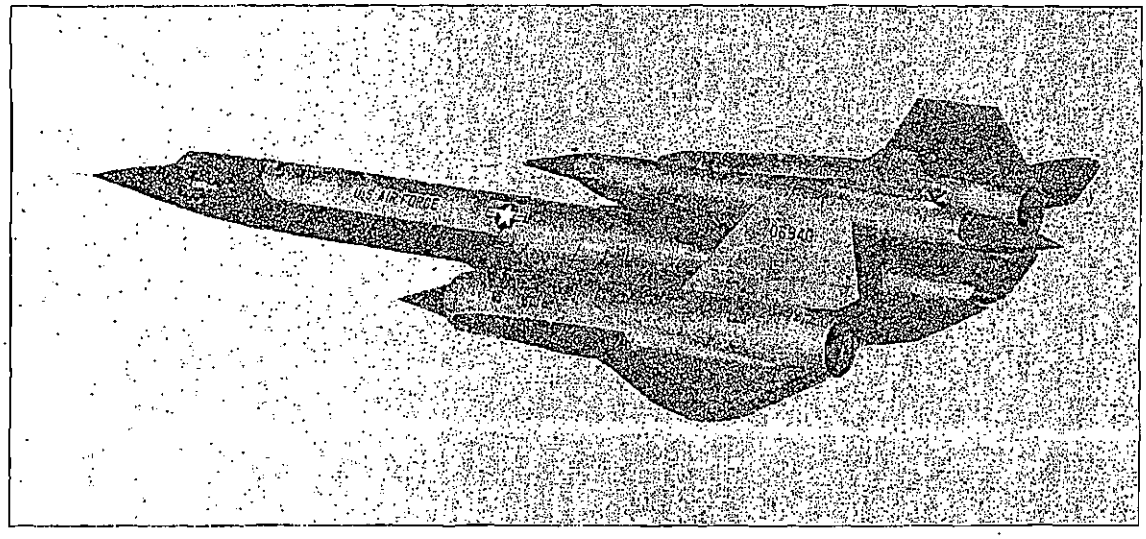
The second new version of the OXCART was another reconnaissance aircraft. In December 1962 the Air Force ordered six "reconnaissance/strike" aircraft, which were designed to conduct high-speed, high-altitude reconnaissance of enemy territory after a nuclear strike. This new aircraft differed from other A-12 versions in that it was longer, had a full-blown two-seat cockpit, and carried a large variety of photographic and electronic sensors. The additional weight of all this equipment gave the Air Force craft a slower maximum speed and a lower operating ceiling than the Agency's A-12. In August 1963, the Air Force added 25 more aircraft to this contract, for a total of 31.[38]

THE QUESTION OF SURFACING A VERSION OF THE OXCART

As the funds being spent on Air Force versions of the OXCART increased dramatically, the Defense Department became concerned that it could not offer any public explanation for these expenditures. At the same time, Agency and Defense Department officials recognized the growing danger that a crash or sightings of test flights could compromise the program. This led the Defense Department in late 1962 and early 1963 to consider surfacing the Air Force's interceptor version of the A-12 to provide a cover for OXCART sightings or crashes and an explanation for the rise in Air Force spending. Some journalists had also become aware of the aircraft's existence, raising concern that the secret would eventually come out in the press. Agency officials remained reluctant to reveal the existence of any version of the A-12, and the issue soon came to the attention of the PFIAB. James Killian and Edwin Land strongly opposed disclosing OXCART's existence, and in January 1963 they presented their views to President Kennedy at a meeting attended by DCI McCone and Defense Secretary Robert McNamara. Killian, Land, and McCone succeeded in persuading the President and Secretary of Defense to keep the OXCART's existence a secret for the time being.

Later that year supporters of the idea of surfacing the OXCART found a more powerful argument for their proposal—the need to disseminate the supersonic technology that had been developed for the

[38] *OSA History*, chap. 20, pp. 71-72 (~~PS~~ Codeword).



M-12 carrying D-21 Drone

OXCART. This technology would be invaluable for Air Force projects such as the B-70 bomber and for the civilian supersonic transport (SST) then being discussed in Congress. In the fall of 1963, several Presidential advisers expressed their concern to DCI McCone that Lockheed had received a $700 million headstart in the development of supersonic technology, giving the firm a tremendous advantage over other aerospace companies working on a supersonic transport. McCone passed these concerns on to President Kennedy on 12 November 1963, just 10 days before the fateful trip to Dallas. The President instructed CIA and the Defense Department to develop a plan for surfacing the OXCART but to await further discussions with him before taking any action.[39]

President Lyndon B. Johnson received a detailed briefing on the OXCART program from McCone, McNamara, Bundy, and Rusk on 29 November, after just one week in office. McNamara strongly advocated surfacing a version of the OXCART. McCone was more cautious, calling for the preparation of a statement that could be used when surfacing became necessary but arguing that such a step was not

[39] John A. McCone, "Memorandum of Meeting in Cabinet Room for the Purpose of Discussing the Surfacing of the OX," 21 January 1963, DCI records (~~TS~~ Codeword); idem, Memorandum for the Record, Discussion with the President—October 21st—6:00 p.m., 22 October 1963, DCI records (S); *OSA History*, chap. 20, pp. 73-74 (~~TS~~ Codeword).

yet needed. Agreeing with McCone's position, President Johnson said the issue should be reviewed again in February.[40]

One additional argument in favor of surfacing the OXCART was the realization that the aircraft could not be used to fly undetected over the Soviet Union. By 1962 the United States had become aware of the effectiveness of a new Soviet radar system, codenamed TALL KING. The introduction of this computer-controlled radar undercut one of the basic premises of the OXCART program, the assumption that radar operators would not be able to track high-flying supersonic targets visually because of their small, nonpersistent radar returns. By coupling a computer to a radar, the Soviets could now weight the individual radar returns and identify those produced by high-flying, very fast objects.[41]

By February 1964 DCI McCone had become convinced that surfacing was necessary. Soviet development of the TALL KING radar system had eliminated his hope that OXCART would eventually be able to carry out its original intended purpose—overflights of the USSR. The final decision on the issue of surfacing the OXCART came at a National Security Council meeting on 29 February 1964, at which all of the participants supported the decision to surface. That same day President Johnson held a news conference at which he announced the successful development of an "advanced experimental jet aircraft, the A-11, which has been tested in sustained flight at more than 2,000 miles per hour and at altitudes in excess of 70,000 feet."[42]

President Johnson had spoken of the A-11 rather than the Agency's A-12, and the aircraft that was actually revealed to the public was the Air Force's YF-12A interceptor, a project that had already been canceled.[43] Following the President's announcement, two of

---

[40] John A. McCone, "Memorandum for the Record, Meeting with the President, Secretary McNamara, Mr. Bundy and DCI," 29 November 1963, DCI records (~~TS~~); *OSA History*, chap. 20, p. 73 (~~TS~~ Codeword).

[41] *OSA History*, chap. 20, pp. 147-149 (~~TS~~ Codeword).

[42] John A. McCone, Memorandum for the Record, "Discussion at the NSC Meeting, Attended by the President, all members and the four members of the President's personal staff, 29 February 1964," 2 March 1964, DCI records (~~S~~); Minnich, "OXCART Story," p. 14—erroneously identifies the date as 24 February—(~~S~~).

[43] President Johnson's use of the designator A-11 at the press conference has sometimes been called an error, but Kelly Johnson wrote the President's press release and chose this designator for security reasons because it referred to the earlier version of the aircraft that lacked the radar-defeating modifications of the A-12. Johnson, "Archangel log," 25 February 1964.

these aircraft were hastily flown to Edwards Air Force Base. From this point on, the Air Force versions of the OXCART were based at Edwards and provided a diversion so that the faster and higher flying A-12s at the Nevada site could continue testing out of the public eye.

The President's announcement did not mention the CIA's involvement in the project, which remained classified, but keeping the Agency's extensive role in the OXCART a secret was not an easy task. The first step had been to separate the Air Force's versions of the A-12 from the Agency's by moving the Air Force aircraft to California. Next, those firms that were to be given the new technology had to be briefed on the program and agree to abide by the same secrecy agreements then in force with Lockheed. Moreover, everyone witting of OXCART (including those no longer associated with the program, such as Allen Dulles, Richard Bissell, and General Cabell) had been briefed about the impending Presidential announcement, so that they would not think that the need for secrecy about OXCART had ended.[44]

The process of surfacing versions of the OXCART continued on 25 July 1964, when President Johnson revealed the existence of a new Air Force reconnaissance aircraft, which he called the SR-71. Actually, the President was supposed to say RS-71 (for "reconnaissance-strike"). Deciding that renaming the aircraft was easier than correcting President Johnson, the Air Force invented a new category—"strategic reconnaissance"—to explain the SR-71's designation.

ADDITIONAL PROBLEMS DURING FINAL TESTING

The first A-12 crash occurred on 24 May 1963, when a detachment pilot, realizing the airspeed indication was confusing and erroneous, decided to eject. The pilot was unhurt, but the plane was destroyed when it crashed near Wendover, Utah. A cover story for the press described the plane as an F-105. All A-12s were grounded for a week while the accident was investigated. The malfunction was found to be caused by ice that had plugged up a pitot-static tube used to determine airspeed.[45]

[44] *OSA History*, chap. 20, p. 76 (~~OX~~ Codeword).

[45] Ibid., pp. 69-70 (~~OX~~ Codeword).

Two more A-12s were lost in later testing. On 9 July 1964, article 133 crashed while landing when a pitch-control servo device froze, rolling the plane into a wing-down position. Ejecting from an altitude of 120 feet, the pilot was blown sideways out of the craft. Although he was not very high off the ground, his parachute did open and he landed during the parachute's first swing. Fortunately he was unhurt, and no news of the accident filtered out of the base. Eighteen months later, on 28 December 1965, article 126 crashed immediately after takeoff because of an improperly wired stability augmentation system. As in the previous crash, the pilot ejected safely, and there was no publicity connected with the crash. An investigation ordered by DCI McCone determined that the wiring error had resulted from negligence, not sabotage.[46]

The A-12 made its first long-range, high-speed flight on 27 January 1965. The flight lasted 100 minutes, 75 minutes of which were flown at speeds greater than Mach 3.1, and the aircraft covered 2,580 miles at altitudes between 75,600 and 80,000 feet. By this time, the OXCART was performing well. The engine inlet, camera, hydraulic, navigation, and flight-control systems all demonstrated acceptable reliability.

Nevertheless, as the OXCART began flying longer, faster, and higher, new problems arose. The most serious of these problems involved the aircraft's wiring. Continuing malfunctions of the inlet controls, communications equipment, ECM systems, and cockpit instruments were often attributable to wiring failures. Wiring connectors and components had to withstand temperatures above 800°F, structural flexing, vibration, and shock. Such demands were more than the materials could stand. Not all of the OXCART's problems could be traced to materiel failures, however, and Agency officials believed that careless maintenance by Lockheed employees also contributed to malfunctions.[47]

Concerned that Lockheed would not be able to meet the OXCART's schedule for operational readiness, the Office of Special Activities' Director of Technology, John Parangosky, met with Kelly Johnson on 3 August 1965 to discuss the project's problems. Johnson not only assigned more top-level supervisors to the project but also

[46] Ibid., pp. 80-81 (TS Codeword); Minnich, "OXCART Story," pp. 17-18 (S).

[47] *OSA History*, chap. 20, p. 94 (TS Codeword).

decided to go to Nevada and take charge of the OXCART's development himself. His presence made a big difference, as can be seen in his notes in the project log:

> *I uncovered many items of a managerial, materiel and design nature.... I had meetings with vendors to improve their operation.... Changed supervision and had daily talks with them, going over in detail all problems on the aircraft.... Increased the supervision in the electrical group by 500%.... We tightened up the inspection procedures a great deal and made inspection stick.*
>
> *It appears that the problems are one-third due to bum engineering.... The addition of so many systems to the A-12 has greatly complicated the problems, but we did solve the overall problem.*[48]

These improvements in on-site management got the project back on schedule.

By 20 November 1965, the final validation flights for OXCART deployment were finished. During these tests, the OXCART achieved a maximum speed of Mach 3.29, an altitude of 90,000 feet, and sustained flight time above Mach 3.2 of 74 minutes. The maximum endurance test lasted six hours and 20 minutes. On 22 November, Kelly Johnson wrote to Brig. Gen. Jack C. Ledford, head of the Office of Special Activities, stating, "The time has come when the bird should leave its nest."[49]

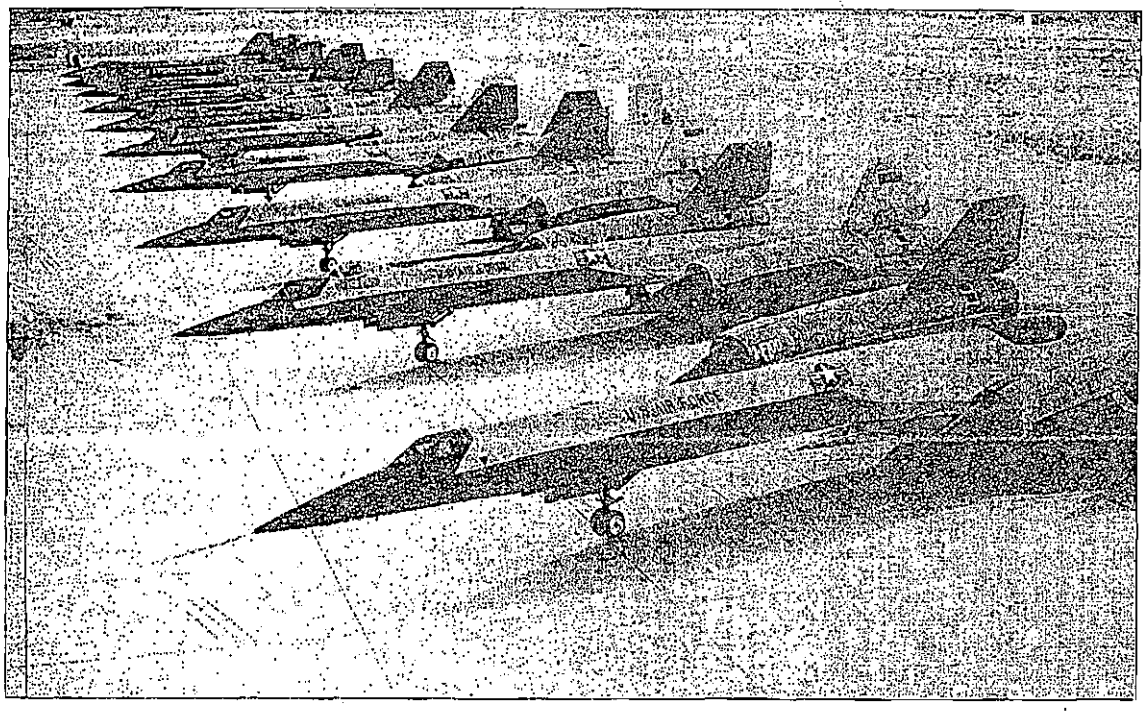
Three years and seven months after its first flight in April 1962, the OXCART was ready for operational use. It was now time to find work for the most advanced aircraft ever conceived and built.

DISCUSSIONS ON THE OXCART'S FUTURE EMPLOYMENT

Although the OXCART had been designed to replace the U-2 as a strategic reconnaissance aircraft to fly over the Soviet Union, this use had become doubtful long before the OXCART was ready for operational use. The U-2 Affair of 1960 made Presidents very reluctant to consider overflights of the Soviet Union. Indeed, Presidents Eisenhower and Kennedy had both stated publicly that the United States would not conduct such overflights. In July 1962, Secretary of

[48] Johnson, "Archangel log," 5 August–30 April 1965.

[49] Minnich, "OXCART Story," p. 23 (S).



A-12s at Area 51

Defense McNamara told DCI McCone that he doubted that the OXCART would ever be used and suggested that improvements in satellite reconnaissance would very likely eliminate the need for the expensive OXCART program. Strongly disagreeing, McCone told McNamara that he had every intention of using OXCART aircraft to fly over the Soviet Union.

McCone raised this issue with President Kennedy in April 1963, at a time when the nation's photosatellites were experiencing a great number of failures and the intelligence community was clamoring for better photography to confirm or disprove allegations of the existence of an antiballistic missile system at Leningrad. Unconvinced by McCone's arguments for OXCART overflights, President Kennedy expressed the hope that some means might be devised for improving satellite imagery instead.[50]

[50] John A. McCone, Memorandum for the Record, "Summary of meeting with Secretary McNamara and Secretary Gilpatric, General Carter and Mr. McCone on 5 July 1962," 6 July 1962, DCI records (~~S~~); McCone, Memorandum for the File, "Meeting with the President—5:30—15 Apr 1963 in Palm Beach, Florida," DCI records (~~S~~).

Although overflights of the Soviet Union appeared to be out of the question, the OXCART's eventual employment elsewhere in the world remained a strong possibility, particularly after the Cuban Missile Crisis of October 1962 demonstrated the continuing need for manned strategic reconnaissance aircraft. Since satellites had not been able to supply the kinds of coverage needed, U-2s had carried out numerous overflights of Cuba. Nevertheless, the U-2 remained vulnerable to surface-to-air missiles (as had once again been demonstrated by the downing of a SAC U-2 during the Missile Crisis), and project headquarters had even briefly considered sending the A-12 over Cuba in October 1962, even though the aircraft still lacked the required J58 engines and would have had to use much less powerful ones.[51] After the Missile Crisis ended, Air Force U-2s continued to photograph Cuba under a tacit superpower understanding that such monitoring of the withdrawal of the missiles would proceed without interference. But the possibility of future Soviet or Cuban action against the U-2s remained, raising the dismaying prospect that the United States would not be able to tell if the Soviet Union was reintroducing ballistic missiles into Cuba.

Such fears became acute in the summer of 1964 after Soviet Premier Nikita Khrushchev told foreign visitors such as columnist Drew Pearson, former Senator William Benton, and Danish Prime Minister Jens Otto Krag that, once the US elections had been held in November, U-2s flying over Cuba would be shot down. Project headquarters therefore began preparing contingency plans (Project SKYLARK) for the possible employment of OXCART over Cuba, even though the new aircraft was not yet ready for operations. On 5 August 1964, the Acting DCI, Gen. Marshall S. Carter, ordered the project staff to achieve emergency operational readiness of the OXCART by 5 November 1964, in case Premier Khrushchev actually carried out his threat to shoot down U-2s.[52]

To meet this deadline, the Office of Special Activities organized a detachment of five pilots and ground crews to conduct flights to validate camera performance and qualify pilots for Mach 2.8 operations. Simulating Cuban missions during training flights, the detachment

[51] On 23 October 1962 Johnson noted in his "Archangel log": that the performance of an A-12 with J75 engines (as suggested by project headquarters for possible use over Cuba) would be "hardly spectacular."

[52] Johnson, "Archangel log," 17 August 1964; Minnich, "OXCART Story," p. 19 (S); *OSA History*, chap. 20, p. 81 (~~TSR~~ Codeword).

demonstrated its ability to conduct overflights of Cuba by the 5 November deadline, which passed without any hostile action by the Soviets or Cubans. The detachment then worked to develop the capability for sustained operations with its five aircraft. All these preparations were valuable training for the OXCART program, even though the SKYLARK contingency plan was never put into effect. Since U-2s continued to satisfy collection requirements for Cuba, the A-12s were reserved for more critical situations.

When the Agency declared that OXCART had achieved emergency operational status on 5 November 1964, the aircraft was still not prepared for electronic warfare, as only one of the several planned electronic countermeasure devices had been installed. Nevertheless, a senior government panel decided that the OXCART could conduct initial overflights of Cuba without a full complement of warning and jamming devices, should the need for such missions arise.

One reason for the delay in completing OXCART's electronic warfare preparations was the Air Force's concern that OXCART use of existing ECM devices could, in the event of the loss of an OXCART over hostile territory, compromise the ECM equipment used by Air Force bombers and fighters. Even if OXCART's ECM devices were merely similar to military ECM systems, the Air Force still worried that their use would give the Soviets an opportunity to work out countermeasures.

Such concerns led the Agency to an entirely different approach to antiradar efforts in Project KEMPSTER. This project attempted to develop electron guns that could be mounted on the OXCART to generate an ion cloud in front of the plane that would reduce its radar cross section. Although this project proved unsuccessful, the CIA also developed a number of more conventional ECM devices for use in the OXCART.[53]

As the OXCART's performance and equipment continued to improve, there was renewed consideration of deploying the aircraft overseas, particularly in Asia, where US military activity was increasing. On 18 March 1965, DCI McCone, Secretary of Defense McNamara, and Deputy Secretary of Defense Vance discussed the

[53] *OSA History*, chap. 20, pp. 149-151 (~~TS~~ Codeword); Notes on the OXCART project by John Parangosky, OSA records (~~TS~~ Codeword).

growing hazards confronting aerial reconnaissance of the People's Republic of China. In three years the Agency had lost four U-2s over China, and the Air Force had lost numerous reconnaissance drones. The three men agreed to go ahead with all the preparatory steps needed for the OXCART to operate over China so that it would be ready in case the President decided to authorize such missions.

Project BLACK SHIELD, the plan for Far East operations, called for OXCART aircraft to be based at Kadena airbase on Okinawa. In the first phase, three planes would be flown to Okinawa for 60-day periods, twice a year, an operation which would involve about 225 personnel. Later there would be a permanent detachment at Kadena. In preparation for the possibility of such operations, the Defense Department spent $3.7 million to provide support facilities and real-time secure communications on the island by early autumn 1965.[54]

In the summer of 1965, after the United States had begun introducing large numbers of troops into South Vietnam, Southeast Asia became another possible target for the OXCART. Because the continued use of U-2s for reconnaissance missions over North Vietnam was threatened by the deployment of Soviet-made surface-to-air missiles, McNamara asked the CIA on 3 June 1965 whether it would be possible to substitute OXCART aircraft for U-2s. The new DCI, Adm. William F. Raborn, replied that the OXCART could operate over Vietnam as soon as it had passed its final operational readiness tests.[55]

Formal consideration of proposed OXCART missions involved the same approval process that was used for U-2 overflights. In late November 1965, after the OXCART had passed its final validation tests, the 303 Committee met to consider a proposal to deploy the OXCART to Okinawa to overfly Southeast Asia and China. Although the committee did not approve deployment, it ordered the development and maintenance of a quick-reaction capability, ready to deploy to Okinawa within 21 days after notification.

There the matter remained for more than a year. During the first half of 1966, DCI Raborn raised the issue of deploying the OXCART to Okinawa at five separate 303 Committee meetings but failed to win

[54] *OSA History*, chap. 20, pp. 90-91 (~~TS~~ Codeword).

[55] Minnich, "OXCART Story," p. 21 (~~S~~).

sufficient support. The JCS and the PFIAB supported the CIA's advocacy of OXCART deployment. Top State and Defense Department officials, however, thought that the political risks of basing the aircraft in Okinawa—which would almost certainly disclose it to the Japanese—outweighed any gains from the intelligence the OXCART might gather. On 12 August 1966, the divergent views were presented to President Johnson, who upheld the 303 Committee's majority opinion against deployment for the time being.[56]

The CIA then proposed an OXCART overflight of Cuba in order to test the aircraft's ECM systems in a hostile environment. On 15 September the 303 Committee considered and rejected this idea on the grounds that sending OXCART over Cuba "would disturb the existing calm prevailing in that area of our foreign affairs." [57]

With operational missions still ruled out, proficiency training remained the main order of business. This led to improvements in mission plans and flight tactics that enabled the detachment to reduce the time required to deploy to Okinawa from 21 days to 15. Records continued to fall to the OXCART. On 21 December 1966, a Lockheed test pilot flew an A-12 for 16,408 kilometers over the continental United States in slightly more than six hours, for an average speed of 2,670 kilometers per hour (which included in-flight refueling at speeds as low as 970 kilometers per hour). This flight set a record for speed and distance unapproachable by any other aircraft.[58]

Two weeks later, on 5 January 1967, an A-12 crashed after a fuel gauge malfunctioned and the aircraft ran out of fuel short of the runway. Pilot Walter Ray ejected but was killed, when he could not become separated from the ejection seat. To preserve the secrecy of the OXCART program, the Air Force informed the press that an SR-71 was missing and presumed down in Nevada. This loss, like the three preceding crashes, did not result from difficulties caused by high-speed, high-temperature flight but from traditional problems inherent in any new aircraft.

Proposals for OXCART operations continued to surface, and in May 1967 the CIA forwarded a detailed request to the 303 Committee to use the OXCART to collect strategic intelligence about a new

---

[56] Minnich, "OXCART Story," p. 23 (~~S~~); *OSA History*, chap. 20, pp. 110-111 (~~TS~~ Codeword).

[57] *OSA History*, chap. 20, p. 112 (~~TS~~ Codeword).

[58] Minnich, "OXCART Story," p. 24 (~~S~~).

Soviet missile system. As early as 1962, the intelligence community began to be concerned about the actual purpose of new missile installations that first appeared near Tallinn, Estonia, and soon spread along the northwestern quadrant of the Soviet Union. Attempts to photograph the sites using reconnaissance satellites had been frustrated by the prevailing cloud cover in the region. Because of the lack of accurate information about the missile sites, there was a wide divergence of views within the intelligence community about their purpose. These views ranged from the CIA's belief that the installations contained long-range, surface-to-air missiles designed to counter strategic bombers, to the Air Force's contention that Tallinn sites represented a deployed antiballistic missile system.

Photointerpreters insisted that imagery with a resolution of 12 to 18 inches was necessary to determine missile size, antenna pattern, and configuration of the engagement radars associated with the system. Electronic intelligence (ELINT) analysts also needed data about the Tallinn radars, but there were no collection sites that could monitor the Tallinn emanations when the radars were being tested. Moreover, the Soviets never operated the radars in the tracking and lockon modes, a fact that prevented analysts from knowing the frequencies or any other performance characteristics of the radar.

To settle the question of the purpose of the Tallinn installations, Office of Special Activities planners proposed a mission that would use the high resolution of the OXCART's camera along with the U-2's sophisticated ELINT-collection equipment. This project's unclassified name was Project SCOPE LOGIC; its classified title was Operation UPWIND.

The proposed project involved launching an A-12 OXCART aircraft from Area 51 in Nevada and flying it to a Baltic Sea rendezvous with a Project IDEALIST U-2 flying from an RAF facility in Great Britain. The OXCART would fly north of Norway and then turn south along the Soviet-Finnish border. Shortly before Leningrad, the A-12 would head west-southwest down the Baltic Sea, skirting the coasts of Estonia, Latvia, Lithuania, Poland, and East Germany before heading west to return to Area 51. The entire flight would cover 11,000 nautical miles, take eight hours and 38 minutes, and require four aerial refuelings.

Although the A-12 would not violate Soviet airspace during this dash, it would appear to Soviet radar network operators to be headed for an overflight penetration in the vicinity of Leningrad. It was

hoped that the A-12's passage would provoke Soviet air defense personnel to activate the Tallinn system radars in order to track the swift OXCART aircraft. As the A-12 made its dash down the Baltic, its Type-I camera would be filming the entire south coast. If Agency analysts were correct in their assumption that the Tallinn system was designed to counter high-altitude aircraft at long ranges, then the OXCART would be in jeopardy during this dash down the Baltic. Nevertheless, Agency weapons experts believed that the A-12 aircraft's speed and suite of electronic countermeasures would keep it safe from the standard Soviet surface-to-air missile installations.

While the A-12 was conducting its high-speed dash along the Baltic coast of Eastern Europe, the U-2 would be flying farther out to sea, safely beyond the range of all Soviet SAMs. The U-2 would be able to collect the Tallinn radar installation's ELINT emanations.

Agency and Defense Department officials supported the proposed mission, but Secretary of State Dean Rusk strongly opposed it and the 303 Committee never forwarded the proposal to President Johnson.[59] The Tallinn radar installation remained of great interest to the intelligence community, and in the late 1960s the CIA attempted to develop a small, unmanned reconnaissance aircraft that could photograph Tallinn and other coastal areas. The project (AQUILINE) was abandoned in 1971 (see appendix E).

FIRST A-12 DEPLOYMENT: OPERATION BLACK SHIELD

Although the Tallinn mission was still being considered in May 1967, another possible employment for the OXCART came under discussion. This time the proposal was for OXCART to collect tactical rather than strategic intelligence. The cause was apprehension in Washington about the possible undetected introduction of surface-to-surface missiles into North Vietnam. When President Johnson asked for a proposal on the matter, the CIA suggested that the OXCART be used. While the State and Defense Departments were still examining the proposal's political risks, DCI Richard Helms

[59] Memorandum for DDCI R. L. Taylor from C. E. Duckett, DDS&T, "Collection of Photo and ELINT Data on Tallinn Sites Utilizing the OXCART and the U-2," 4 May 1967, DS&T records (~~PS~~ Codeword).

raised the issue at President Johnson's "Tuesday lunch" on 16 May. Helms got the President's approval, and the CIA put the BLACK SHIELD plan to deploy the OXCART to the Far East into effect later that same day.[60]

The airlift of personnel and equipment to Kadena began on 17 May 1967, and on 22 May the first A-12 flew nonstop from Area 51 to Kadena in six hours and six minutes. A second aircraft arrived on 24 May. The third A-12 left on 26 May, but the pilot had trouble with the inertial navigation system and communications near Wake Island. He made a precautionary landing at Wake, where a pre-positioned emergency recovery team was located. The problem was corrected and the aircraft continued its flight to Kadena on the following day.

Before the start of the operation, the CIA briefed a number of key US and Allied officials on the operation. Included were the US Ambassadors

25X1, E.O.13526

the Prime Ministers of Japan and Thailand; the President and Defense Minister of the Republic of China; and the Air Force chiefs of Thailand and the Republic of China. All of these officials favored the operation.

By 29 May 1967, 13 days after President Johnson's approval, BLACK SHIELD was ready to fly an operational mission. On 30 May, the detachment was alerted for a mission on the following day. As the takeoff time approached, Kadena was being deluged by rain, but, since weather over the target area was clear, flight preparations continued. The OXCART, which had never operated in heavy rain, taxied to the runway and took off.

This first BLACK SHIELD mission flew one flight path over North Vietnam and another over the demilitarized zone (DMZ). The mission was flown at Mach 3.1 and 80,000 feet and lasted three hours and 39 minutes. While over North Vietnam, the A-12 photographed 70 of the 190 known surface-to-air missile sites and nine other priority targets. The A-12's ECM equipment did not detect any radar signals during the mission, which indicated that the flight had gone completely unnoticed by both the Chinese and North Vietnamese.

[60] Minnich, "OXCART Story," p. 25 (S).

During the next six weeks, there were alerts for 15 BLACK SHIELD missions, seven of which were actually flown. Only four detected hostile radar signals. By mid-July 1967, the BLACK SHIELD missions had provided sufficient evidence for analysts to conclude that no surface-to-surface missiles had been deployed in North Vietnam.[61]

Project Headquarters in Langley planned and directed all operational BLACK SHIELD missions. To ensure secure communications between Headquarters and Kadena,

25X1, E.O.13526

A typical mission over North Vietnam required refueling south of Okinawa, shortly after takeoff. After the planned photographic passes, the aircraft withdrew for a second aerial refueling in the Thailand area before returning to Kadena. So great was the plane's speed that it spent only 12.5 minutes over Vietnam during a "single-pass" mission, and 21.5 minutes during a "two-pass" mission. Because of its wide 86-mile turning radius, the plane occasionally crossed into Chinese airspace when getting into position for a second pass.

After the aircraft landed, the camera film was removed and sent by special plane to processing facilities in the United States. By late summer, however, an Air Force photo laboratory in Japan began doing the processing in order to place the photointelligence in the hands of US commanders in Vietnam within 24 hours of a mission's completion.

BLACK SHIELD activity continued unabated during the second half of 1967. From 16 August to 31 December 1967, 26 missions were alerted and 15 were flown. On 17 September one SAM site tracked the vehicle with its acquisition radar but was unsuccessful with its FAN SONG guidance radar. It was not until 28 October that a North Vietnamese SAM site launched a missile at the OXCART. Mission photography documented the event with photographs of missile smoke above the SAM firing site and pictures of the missile and its contrail. Electronic countermeasures equipment aboard the OXCART performed well, and the missile did not endanger the aircraft.

[61] Minnich, "OXCART Story," pp. 25-28 (~~S~~); *OSA History*, chap. 20, pp. 119-124, annex 152 (~~TS~~ Codeword).

The only time the enemy came close to downing an OXCART was on 30 October 1967. During his first pass over North Vietnam, pilot Dennis Sullivan detected radar tracking. Two SAM sites prepared to launch missiles but neither did. During Sullivan's second pass the North Vietnamese fired at least six missiles at the OXCART, each confirmed by vapor trails on mission photography. The pilot saw these vapor trails and witnessed three missile detonations near but behind the A-12, which was traveling at Mach 3.1 at about 84,000 feet. Postflight inspection of the aircraft revealed that a piece of metal had penetrated the underside of the right wing, passed through three layers of titanium, and lodged against a support structure of the wing tank. The fragment was not a warhead pellet but probably debris from one of the missile detonations that the pilot observed.[62]

BLACK SHIELD missions continued during the first three months of 1968, with four missions flown over North Vietnam out of 14 alerts. The last OXCART overflight of Vietnam took place on 8 March 1968. During this same three-month period, the OXCART made its first overflight of North Korea after the USS Pueblo was seized on 23 January 1968. The goal of this mission was to discover whether the North Koreans were preparing any large-scale hostile move in the wake of this incident. When NPIC photointerpreters examined OXCART photography taken on 26 January, they found the missing USS Pueblo in Wonsan harbor.

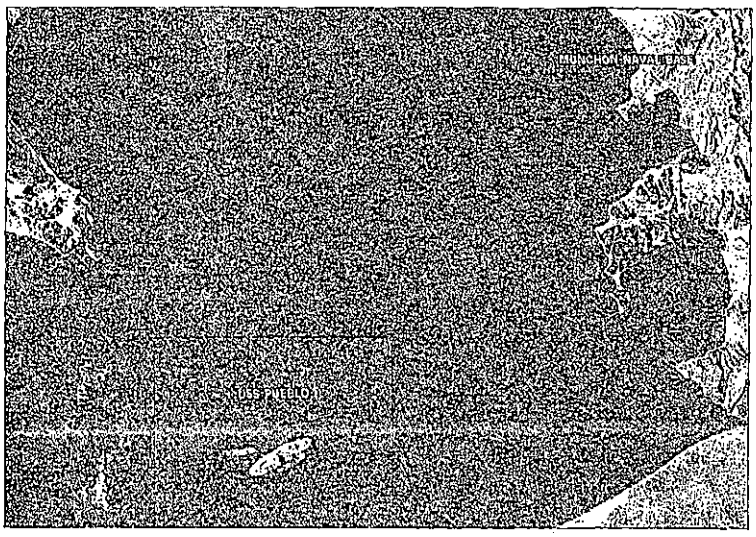
Secretary of State Dean Rusk was reluctant to endorse a second mission over North Korea for fear of diplomatic repercussions should the aircraft come down in hostile territory. The Secretary was assured that the plane could transit North Korea in seven minutes and was unlikely to land in either North Korea or China. The 303 Committee then endorsed a second mission over North Korea, which was flown on 19 February. A third and final overflight of North Korea on 8 May 1968 proved to be the last operational deployment of the OXCART aircraft.[63]

THE END OF THE OXCART PROGRAM

Almost a decade had elapsed between the time when the concept for the OXCART aircraft was first examined and the first A-12 was operationally deployed. Now after only 29 operational missions, the most

[62] Minnich, "OXCART Story," p. 28 ~~(S)~~.

[63] Ibid., pp. 28-29 ~~(S)~~.



USS Pueblo in Wonsan Harbor

advanced aircraft ever built was to be put out to pasture. The abandonment of the OXCART did not result from any shortcomings of the aircraft; the causes lay in fiscal pressures and competition between the reconnaissance programs of the CIA and the Air Force.

Throughout the OXCART program, the Air Force had been exceedingly helpful; it gave financial support, conducted the refueling program, provided operational facilities at Kadena, and airlifted OXCART personnel and supplies to Okinawa for the Vietnam and Korean operations. Air Force orders for variants of the CIA's A-12—the YF-12A interceptor and the SR-71 reconnaissance aircraft—had helped lower development and procurement costs for the OXCART. Nevertheless, once the Air Force had built up its own fleet of reconnaissance aircraft, budgetary experts began to criticize the existence of two expensive fleets of similar aircraft.

In November 1965, the very month that the A-12 had been declared operational, the Bureau of the Budget circulated a memorandum that expressed concern about the costs of the A-12 and SR-71 programs. It questioned both the total number of planes required for the combined fleets, and the necessity for a separate CIA fleet. The memorandum recommended phasing out the A-12 program by September 1966 and stopping any further procurement of the SR-71

models. The Secretary of Defense rejected this recommendation, presumably because the SR-71 would not be operational by September 1966.[64]

In July 1966, at the Bureau of the Budget's suggestion, a study group was established to look for ways to reduce the cost of the OXCART and SR-71 programs. The study group consisted of C. W. Fischer from the Bureau of the Budget, Herbert Bennington from the Department of Defense, and John Parangosky from CIA. The study group listed three possible courses of action: maintain both fleets, mothball the A-12s but share the SR-71s between CIA and the Air Force, or mothball the A-12s and assign all missions to Air Force SR-71s. On 12 December 1966, four high-level officials met to consider these alternatives. Over the objections of DCI Helms, the other three officials—Deputy Secretary of Defense Cyrus Vance, Bureau of the Budget Director Charles L. Schultze, and Presidential Scientific Adviser Donald F. Hornig—decided to terminate the OXCART fleet. Concerned that this recommendation would strip the CIA of its supersonic reconnaissance capability, Helms then asked that the SR-71 fleet be shared between CIA and the Air Force.[65]

Four days later, Schultze handed Helms a draft memorandum for the President requesting a decision either to share the SR-71 fleet between CIA and the Air Force or to terminate the CIA capability entirely. Having just received new information indicating that the SR-71's performance was inferior to that of the A-12, Helms asked for another meeting to review this data. His concern was that the SR-71 could not match the photographic coverage that the A-12 could provide. Only one of the SR-71's three camera systems was working anywhere near the original specifications, and that was its Operational Objective system which could only photograph a swath 28 miles wide with a resolution of 28 to 30 inches. The A-12's Type-I P-E camera could photograph a swath 72 miles wide with a nadir resolution of 12 to 18 inches and oblique resolution of 54 inches. Thus, the A-12's camera covered three times as much territory as the SR-71's camera and did so with better resolution. In addition, the A-12 could fly 2,000 to 5,000 feet higher than the SR-71 and was also faster, with a maximum speed of Mach 3.1 compared with the SR-71's Mach 3.0.[66]

---

[64] *OSA History*, chap. 20, p. 130 (~~TS~~ Codeword); Minnich, "OXCART Story," p. 30 ~~(S)~~.

[65] *OSA History*, chap. 20, pp. 130-133 (~~TS~~ Codeword); Minnich, "OXCART Story," pp. 30-31 ~~(S)~~.

[66] Minnich, "OXCART Story," p. 31 (S); *OSA History*, pp. 133-134 (~~TS~~ Codeword).

In spite of Helms's request and the strength of his arguments, the Bureau of the Budget memorandum was submitted to President Johnson. On 28 December 1966, the President approved the termination of the OXCART program by 1 January 1968.

This decision meant that CIA had to develop a schedule for an orderly phaseout of the A-12. This activity was known as Project SCOPE COTTON. Project headquarters informed Deputy Defense Secretary Vance on 10 January 1967 that the A-12s would gradually be placed in storage, with the process to be completed by the end of January 1968. In May 1967, Vance directed that SR-71s would assume responsibility for Cuban overflights by 1 July 1967 and would add responsibility for overflights of Southeast Asia by 1 December 1967. Until these capabilities were developed, OXCART was to remain able to conduct assignments on a 15-day notice for Southeast Asia and a seven-day notice for Cuba.[67]

All these arrangements were made before the OXCART had conducted a single operational mission, which did not occur until 31 May 1967. In the months that followed the initiation of operations in Asia, the OXCART demonstrated its exceptional technical capabilities. Soon some high-level Presidential advisers and Congressional leaders began to question the decision to phase out OXCART, and the issue was reopened.

The CIA contended that the A-12 was the better craft because it flew higher, faster, and had superior cameras. The Air Force maintained that its two-seat SR-71 had a better suite of sensors, with three different cameras (area search, spotting, and mapping), infrared detectors, side-looking aerial radar, and ELINT-collection gear. In an effort to resolve this argument, the two aircraft were pitted against each other in a flyoff codenamed NICE GIRL. On 3 November 1967, an A-12 and an SR-71 flew identical flight paths, separated in time by one hour, from north to south roughly above the Mississippi River. The data collected during these missions were evaluated by representatives of the CIA, DIA, and other Defense Department intelligence organizations.

The results proved inconclusive. Both photographic systems provided imagery of sufficient quality for analysis. The A-12 Type-I camera's 72-mile swath width and 5,000-foot film supply were superior to the SR-71 Operational Objective camera's 28-mile swath and

[67] Minnich, "OXCART Story," p. 31 (S); *OSA History*, p. 138 (~~TS~~ Codeword).

3,300-foot film supply. On the other hand, the SR-71's infrared, side-looking aerial radar, and ELINT/COMINT equipment provided some unique intelligence not available from the A-12. Air Force planners admitted, however, that some of this equipment would have to be sacrificed in order to provide the SR-71 with ECM gear.[68]

Although the flyoff had not settled the question of which aircraft was superior, the OXCART did win a temporary reprieve in late November 1967. The Johnson administration decided to keep both fleets for the time being, particularly because the OXCART was actually flying missions over North Vietnam. With expenditures for the Vietnam war rising steadily, the question of reducing the costs of competing reconnaissance programs was bound to surface again. In the spring of 1968, there was yet another study of the OXCART and SR-71 programs. On 16 May 1968, the new Secretary of Defense, Clark Clifford, reaffirmed the original decision to terminate the OXCART program and store the aircraft. President Johnson confirmed this decision on 21 May.[69]

Project headquarters selected 8 June 1968 as the earliest possible date for phasing out all OXCART aircraft. Those A-12s already at the Nevada site were placed in storage, and the aircraft on Okinawa were scheduled to return by 8 June. Unfortunately, tragedy struck before this redeployment took place. On 4 June 1968 during a test flight from Kadena to check out a new engine, an A-12 disappeared 520 miles east of Manila. Search and rescue missions found no trace of the plane or its pilot, Jack W. Weeks. Several days later the remaining two A-12s left Okinawa to join the other eight OXCART aircraft in storage at Palmdale, California. Because the A-12s were smaller than either of the Air Force's versions, the only parts that could be salvaged for Air Force use were the J58 engines. The OXCART's outstanding Perkin-Elmer camera cannot be used in the SR-71 because the two-seater Air Force aircraft has a smaller camera compartment than that of the A-12. Constructed from one of the most durable metals known to man but unable to fly for want of engines, the OXCART aircraft are fated to remain inactive at Palmdale for many, many years.

---

[68] Information supplied by James Cunningham to Donald E. Welzenbach.

[69] Minnich, "OXCART Story," pp. 32-33 (S); *OSA History*, chap. 20, pp. 143-146 (TS Codeword).

Initial storage arrangements for A-12s at Palmdale

POSSIBLE SUCCESSORS TO THE OXCART

The OXCART was the last high-altitude reconnaissance aircraft produced for the CIA, although the Office of Special Activities did briefly consider several possible successors to the OXCART during the mid-1960s. The first of these, known as Project ISINGLASS, was prepared by General Dynamics to utilize technology developed for its Convair Division's earlier FISH proposal and its new F-111 fighter in order to create an aircraft capable of Mach 4-5 at 100,000 feet. General Dynamics completed its feasibility study in the fall of 1964, and OSA took no further action because the proposed aircraft would still be vulnerable to existing Soviet countermeasures. In 1965 a more ambitious design from McDonnell Aircraft came under consideration as Project RHEINBERRY (although some of the work seems to have come under the ISINGLASS designation as well). This proposal featured a rocket-powered aircraft that would be launched from a B-52 mother ship and ultimately reach speeds as high as Mach 20 and altitudes of up to 200,000 feet. Because building this aircraft would have involved tremendous technical challenges and correspondingly high costs, the Agency was not willing to embark on such a program at a time when the main emphasis in overhead reconnaissance had shifted from aircraft to satellites. As a result, when the OXCART program ended in the summer of 1968, no more advanced successor was waiting in the wings—only the veteran U-2.

SUMMARY OF THE OXCART PROGRAM

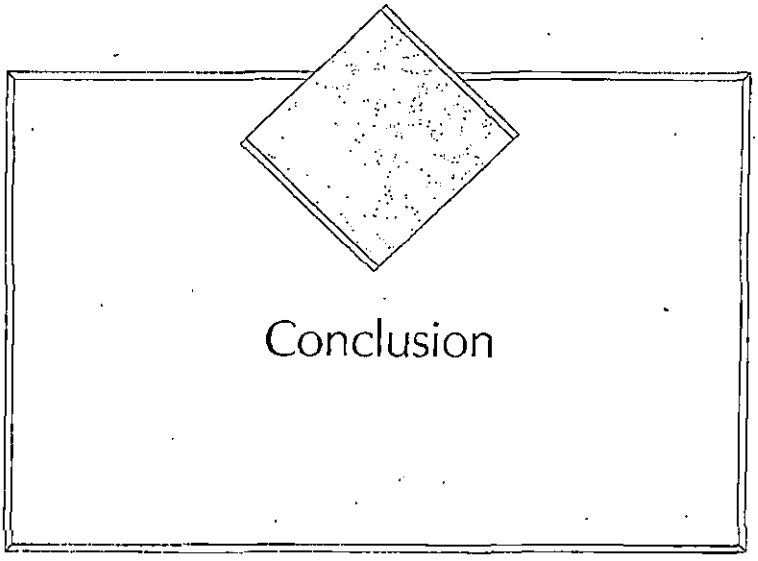
Intended to replace the U-2 as a collector of strategic intelligence, the OXCART was never used for this purpose. Its brief deployment was strictly for obtaining tactical intelligence and its photographic product contributed very little to the Agency's strategic intelligence mission. By the time OXCART became operational, photosatellite systems had filled the role originally conceived for it. The most advanced aircraft of the 20th century had become an anachronism before it was ever used operationally.[20]

The OXCART did not even outlast the U-2, the aircraft it was supposed to replace. The OXCART lacked the quick-response capability of the smaller craft: a U-2 unit could be activated overnight, and within a week it could deploy abroad, fly sorties, and return to home base. The OXCART planes required precise logistic planning for fuel and emergency landing fields, and their inertial guidance systems needed several days for programming and stabilization. Aerial tankers had to be deployed in advance along an OXCART's flight route and be provisioned with the highly specialized fuel used by the J58 engines. All of this required a great deal of time and the effort of several hundred people. A U-2 mission could be planned and flown with a third fewer personnel.

Although the OXCART program created a strategic reconnaissance aircraft with unprecedented speed, range, and altitude, the program's most important contributions lay in other areas: aerodynamic design, high-impact plastics, engine performance, cameras, electronic countermeasures, pilot life-support systems, antiradar devices, use of nonmetallic materials for major aircraft assemblies, and improvements in milling, machining, and shaping titanium. In all of these areas, the OXCART pushed back the frontiers of aerospace technology and helped lay the foundation for future "stealth" research.

---

[20] On 26 January 1967 Kelly Johnson noted in his "Archangel log":
I think back to 1959, before we started this airplane, to discussions with Dick Bissell where we seriously considered the problem of whether there would be one more round of aircraft before the satellites took over. We jointly agreed there would be just one round, and not two. That seems to have been a very accurate evaluation, as it seems that 30 SR-71's give us enough overflight reconnaissance capability and we don't need the additional 10 Oxcart aircraft.



Conclusion

U-2 OVERFLIGHTS OF THE SOVIET UNION

Before the first U-2 overflights in the summer of 1956, project managers believed that their aircraft could fly virtually undetected over the Soviet Union. They did not expect this advantage to last very long, however, because they also expected the Soviets to develop effective countermeasures against the U-2 within 12 to 18 months. Recognizing that time was against them, the U-2 project managers planned a large number of missions to obtain complete coverage of the Soviet Union as quickly as possible. At this time, the U-2 program focused solely on the collection of strategic intelligence.

Once operations began, however, project managers found themselves operating under severe constraints. Contrary to the CIA's expectations, the U-2 could not fly undetected. Its overflights led to Soviet diplomatic protests and numerous attempts at interception. Not wishing to aggravate the Soviet Union during periods of tension or to harm relations during more favorable intervals, President Eisenhower placed strict limits on overflights, personally authorizing each one and greatly limiting their number. Yet, the President never went so far as to eliminate the overflight program. As Commander in Chief, he valued the intelligence that the U-2 overflights collected, especially at times when the press and Congress alleged that the United States was falling behind the Soviet Union militarily, first in bombers and then in missiles. As a result of the President's ambivalence toward overflights, the years 1956-60 were marked by long periods during which no overflights occurred, followed by brief bursts of activity.

The low level of overflight activity did not prevent the U-2 from accomplishing a lot in the four years it flew over the Soviet Union and Eastern Europe. Twenty-four U-2 missions made deep penetration overflights of the Soviet Union: six by Detachment A from Germany, three by Detachment C from the Far East and Alaska, and 15 by Detachment B from Turkey and Pakistan, including the unsuccessful Powers mission.

The amount of information these missions gathered was impressive. By the summer of 1960, the U-2 project had developed more than 1, 285,000 feet of film—a strip almost 250 miles long. The U-2s covered more than 1,300,000 square miles of the Soviet Union, approximately 15 percent of its total area. Information from U-2 photographs was used to prepare 5,425 separate photoanalytical reports.[1]

Numbers alone cannot describe the importance of the U-2 overflight project. In a 28 May 1960 memorandum, after Powers was shot down, DCI Allen W. Dulles described the program's accomplishments: "Five years ago, before the beginning of the U-2 program, . . . half knowledge of the Soviet Union and uncertainty of its true power position posed tremendous problems for the United States. We were faced with the constant risk of exposing ourselves to enemy attack or of needlessly expending a great deal of money and effort on misdirected military preparations of our own." Dulles went on to describe the U-2's contribution in gathering information on four critical aspects of the Soviet Union's power position: its bomber force, its missile force, its atomic energy program, and its air defense system.[2]

The first major contribution of intelligence collected from U-2 overflights was the exposure of the "bomber gap" as a myth. Contrary to the US Air Force's claims, the Soviet Union was not building a large force of long-range bombers. Armed with information from U-2 overflights, President Eisenhower was able to resist pressure to build a large US bomber fleet to meet a nonexistent Soviet threat.

---

[1] DCI Allen W. Dulles, Memorandum for Brig. Gen. Andrew J. Goodpaster, "Statistics Relating to the U-2 Program," 19 August 1960. Operation MUDLARK files, OSA records, job 74-B-605, box 2 (~~CPS~~ Codeword).

[2] The original draft of this document was probably written by James Q. Reber. It was then revised by DCI Dulles. "Accomplishments of the U-2 Program," 27 May 1960, Operation MUDLARK files, OSA records, job 74-B-605, box 2 (~~CPS~~ Codeword).

The "bomber-gap" controversy was soon followed by a "missile-gap" controversy, provoked by an extensive Soviet propaganda campaign that claimed a substantial Soviet lead in developing and deploying ICBMs. U-2 missions searched huge stretches of the Soviet Union along the rail network, looking for ICBMs deployed outside the known missile testing facilities. These missions enabled the CIA to conclude, as Dulles explained to Congress in May 1960, that "the Soviet ICBM program has not been and is not now a *crash* program; instead, it is an orderly, well-planned, high-priority program aimed at achieving an early ICBM operational capability."[3] As with the controversy over Soviet bomber strength, information from U-2 photography enabled President Eisenhower to resist pressure to accelerate the US missile deployment program by building obsolescent liquid-fueled missiles rather than waiting to complete the development of more reliable solid-fueled missiles.

U-2 missions also gathered considerable data on the Soviet Union's atomic energy program, including the production of fissionable materials, weapons development and testing activities, and the location and size of nuclear weapons stockpile sites. Such U-2 photography also revealed no evidence that the Soviet Union had violated the nuclear testing moratorium.

One of the greatest contributions of the U-2 program was to increase the capabilities of the US deterrent force. Before the U-2 overflights, most target information was based on obsolete materials dating back to World War II or shortly thereafter. With the assistance of U-2 photography, the Defense Department could allocate weapons and crews more efficiently and identify many new targets. U-2 photos also proved invaluable in determining the precise location of targets. One further contribution to the capabilities of the US deterrent force was the information that U-2s collected on the Soviet air defense system. U-2 photography located Soviet fighter airfields and gained intelligence on new fighter models. Special electronic intercept and recording equipment carried on many U-2 missions enabled the CIA to analyze the technical characteristics, operational techniques, and radar order of battle of the Soviet Union's electronic defenses. This information was vital both for planning the routes for US deterrent forces and for developing electronic countermeasures.

[3] *Ibid.*, p. 3 (~~CIS~~ Codeword).

The U-2 program not only provided information on individual Soviet weapons systems, but also helped analysts assess basic Soviet intentions, particularly during crisis situations, as Dulles wrote in May 1960:

> *Whenever the international situation becomes tense because of a problem in some particular area, we are concerned whether the situation might get beyond control—that someone on the other side might suddenly and irrationally unleash big war. . . . Our knowledge of Soviet military preparations, however, resulting from the overflight program, has given us an ability to discount or call the bluffs of the Soviets with confidence. We have been able to conclude that Soviet statements were more rhetorical than threatening and that our courses of action could be carried through without serious risk of war and without Soviet interference.*[4]

Dulles closed his report on the U-2's accomplishments by putting the program in perspective as part of the entire national intelligence effort, noting that "in terms of reliability, of precision, of access to otherwise inaccessible installations, its contribution has been unique. And in the opinion of the military, of the scientists and of the senior officials responsible for our national security it has been, to put it simply, invaluable."

The impact of the U-2 overflights on international relations is harder to measure. On the one hand, the intelligence they gathered was a major factor in keeping the United States from beginning a costly and destabilizing arms race in the late 1950s and early 1960s by showing that the Soviet Union was not engaged in major buildups of strategic bombers and intercontinental ballistic missiles. On the other hand, violations of Soviet airspace by U-2s strained relations with Moscow at times and led to the collapse of the 1960 summit meeting. On balance, however, the impact of the U-2 on superpower relations was positive. Without the intelligence gathered by the U-2, the Soviet Union's strategic military capabilities would have remained a mystery, making it very difficult for the President to resist pressure from the military, the Congress, and the public to carry out major increases in strategic weapons, which would have poisoned relations with the Soviet Union far more than the small number of overflights did.

---

[4] Ibid., pp. 9-10 (~~CPS~~ Codeword).

PARTICIPATION OF ALLIES IN THE U-2 PROGRAM

From the very beginning of the overflight project, US Allies provided valuable support. Bases in Germany, Turkey, Norway, and Pakistan played a major role in overflights of the Soviet Union. Bases in India, Thailand, Japan, and the Philippines greatly assisted operations in Asia. Two Allies—the United Kingdom and Nationalist China—made an even greater contribution to the U-2 program by providing pilots and conducting overflights. British pilots began flying in late 1958 and conducted two important overflights of the Soviet Union in late 1959 and early 1960. After the end of such missions in May 1960, the need for British participation lessened. RAF pilots henceforth flew only training or ferry missions, although their use for operational missions was considered on several occasions.

The end of overflights of the Soviet Union reduced the importance of British participation but resulted in the addition of a new source of pilots when the focus of interest for the U-2 in its strategic-intelligence-gathering role shifted to the People's Republic of China. The United States and Nationalist China had been conducting joint reconnaissance projects over the Chinese mainland since the mid-1950s, and in 1961 the CIA equipped the Nationalist Chinese with the latest in reconnaissance aircraft, the U-2. For the next 12 years, Agency U-2s with Nationalist Chinese pilots brought back large quantities of information on the development of Communist China's armed forces, nuclear technology, and economy. Such information was extremely important to US policymakers. Nationalist China paid a high price in lives for its participation in the U-2 program: seven pilots died (five in training accidents and two on overflights), and another three were captured.

U-2s AS COLLECTORS OF TACTICAL INTELLIGENCE

The low level of mission activity over the project's original target—the Soviet Union—was initially very frustrating for CIA project managers, but the U-2 soon found new missions not originally envisioned for the program. With its strategic-intelligence-collection role often on hold, the U-2 became highly useful as a collector of tactical intelligence during crisis situations.

Beginning with the Suez Crisis of 1956 and continuing with subsequent Middle Eastern wars, a rebellion in Indonesia, Sino-Indian border fighting, and culminating in support to the growing US

involvement in Indochina, U-2 photography provided accurate and up-to-date intelligence to US policymakers and field commanders, assisting them in crisis management and the planning of military operations. Agency U-2s also assisted in monitoring cease-fire agreements in the Middle East, with operations occurring after an undeclared war in 1970 and the 1973 Middle East war.

By the time the OXCART became fully operational, manned strategic reconnaissance of the Soviet Union was no longer seriously considered. The political risks were too high, especially since the quality of intelligence from reconnaissance satellites was increasing steadily. Thus, the OXCART's only operational use was for collecting tactical intelligence in the Far East. Like the U-2, the OXCART gathered valuable intelligence during crisis situations. Thus, in January 1968, OXCART photography revealed the location of the USS Pueblo and showed that the North Koreans were not preparing any large-scale military activity in conjunction with the ship's seizure.

ADVANCES IN TECHNOLOGY

One very important byproduct of the CIA's manned reconnaissance program was the many advances in technology that it generated. Thanks to simplified covert procurement arrangements and the lack of detailed and restricting specifications, creative designers such as Kelly Johnson produced state-of-the-art aircraft in record time. The U-2, designed to carry out reconnaissance missions for two years at best, proved so successful that, even after its original area of activity became too dangerous for overflights at the end of four years, the aircraft served the CIA well for another 14 years and still is in service with other government agencies.

The OXCART is an even better example of the technological advances generated by the CIA's reconnaissance program. Although the OXCART was designed almost 30 years ago and first flown in 1962, its speed and altitude have never been equaled. The development of this aircraft also led to the use of new materials in aircraft construction. Unfortunately, the technological breakthroughs that made the OXCART possible took longer than expected. By the time the aircraft was ready for operations, the missions originally planned for it were not practicable. The tremendous technological achievement represented by the OXCART ultimately led to the aircraft's demise by inspiring the Air Force to purchase its own version of the aircraft. The

government could not afford to maintain two such similar reconnaissance programs. The elimination of the Agency's OXCART program did not, however, spell the end of the usefulness of the world's most advanced aircraft; its offspring, the SR-71, is still in service.

In addition to the aircraft themselves, many other items associated with the reconnaissance program have represented important advances in technology. The flight suits and life-support systems of the U-2 and OXCART pilots were the forerunners of the equipment used in the space program. Camera resolution improved dramatically as the result of cameras and lenses produced for the CIA's reconnaissance program.

COOPERATION WITH THE AIR FORCE

In this history, which concentrates on the CIA's involvement in overhead reconnaissance, it is easy to overlook the important role that the US Air Force played in the U-2 and OXCART programs. From the very beginnings of the U-2 program in 1954, the Agency and the Air Force were partners in advancing the state of the art in overhead reconnaissance. Air Force personnel served at all levels of the reconnaissance program, from project headquarters to the testing site and field detachments. The Air Force supplied the U-2's engines, at times diverting them from other high-priority production lines. Perhaps most important of all, the Air Force provided pilots for the U-2s after the Agency's original attempt to recruit a sufficient number of skilled foreign pilots proved unsuccessful. Finally, the day-to-day operations of the U-2s could not have been conducted without the help of Air Force mission planners, weather forecasters, and support personnel in the field detachments. The cooperation between the Agency and the Air Force that began with the U-2 and continued with Project OXCART remains a major feature in US reconnaissance programs today.

IMPACT OF THE OVERHEAD RECONNAISSANCE PROGRAM ON THE CIA

CIA's entry into the world of overhead reconnaissance at the end of 1954 ultimately produced major changes in the Agency. Classical forms of intelligence—the use of covert agents and clandestine operations—gradually lost their primacy to the new scientific and

technical means of collection. As soon as the U-2 began flying over the Soviet Union, its photographs became the most important source of intelligence available. The flood of information that the U-2 missions gathered led to a major expansion of the Agency's photointerpretation capabilities, which finally resulted in the creation of the National Photographic Interpretation Center to serve the entire intelligence community.

The U-2's tremendous success as an intelligence-gathering system led the Agency to search for follow-on systems that could continue to obtain highly reliable information in large quantities. Thus, the CIA sponsored the development of the world's most advanced aircraft—the OXCART—and also pioneered research into photosatellites. Less than a decade after the U-2 program began, the Agency's new emphasis on technical means of collection had brought about the creation of a new science-oriented directorate, which would ultimately rival in manpower and budget the Agency's other three directorates combined.

The negative aspect of this new emphasis on technology is exploding costs. The Agency's first strategic reconnaissance aircraft, the U-2, cost less than $1 million apiece. With the U-2's successor, the OXCART, each aircraft cost more than $20 million, and the cost explosion has continued with each new generation of reconnaissance satellites.

Perhaps the greatest significance of the CIA's entry into the world of overhead reconnaissance in December 1954 was the new national policy that it signaled. Although US military aircraft had frequently violated Soviet airspace in the decade after World War II, such shallow-penetration overflights, concentrating primarily on order-of-battle data, had been authorized and controlled by US field commanders, not by the President. In the autumn of 1954, however, President Dwight D. Eisenhower—determined to avoid another Pearl Harbor—authorized the construction of a new aircraft designed solely to fly over the Soviet Union and gather strategic intelligence. Peacetime reconnaissance flights over the territory of a potential enemy power thus became national policy. Moreover, to reduce the danger of conflict, the President entrusted this mission not to the armed forces, but to a civilian agency—the CIA. From that time forward, overhead reconnaissance has been one of the CIA's most important missions.

APPENDIX A

Acronyms

AEC	Atomic Energy Commission
AFB	Air Force Base
AFDAP	Air Force office symbol for the Assistant for Development Planning under the Deputy Chief of Staff for Development
AMD	Air/Maritime Division
ARC	Ad Hoc Requirements Committee
ARDC	Air Research and Development Command (USAF)
ASPIC	Asian Photographic Interpretation Center
ATIC	Air Technical Intelligence Center (USAF)
BSAP	Boston Scientific Advisory Panel
BUORL	Boston University Optical Research Laboratory
COMINT	Communications Intelligence
COMIREX	Committee on Imagery Requirements and Exploitation
COMOR	Committee on Overhead Reconnaissance
DB	"Dirty Bird"
DCI	Director of Central Intelligence
DCID	Director of Central Intelligence Directive
DDCI	Deputy Director of Central Intelligence
DDI	Deputy Director for Intelligence
DDP	Deputy Director (or Directorate) for Plans
DDS&T	Deputy Director for Science and Technology
DPD	Development Projects Division
DPS	Development Projects Staff
ECM	Electronic Countermeasures
EG&G	Edgerton, Germeshausen & Grier, Incorporated
ELINT	Electronic Intelligence
FCRC	Federally Controlled Research Center
HASP	High-Altitude Air Sampling Program
IAC	Intelligence Advisory Committee
IAS	Indicated air speed
IC	Intelligence community
ICBM	Intercontinental ballistic missile
IR	Infrared
ISP	Intelligence Systems Panel (USAF)
JRC	Joint Reconnaissance Center
MATS	Military Air Transport Service (USAF)

MRBM	Medium-range ballistic missile
NACA	National Advisory Committee for Aeronautics
NAS	Naval air station
NASA	National Aeronautics and Space Administration
NIE	National Intelligence Estimate
NPIC	National Photographic Interpretation Center
NSA	National Security Agency
NSC	National Security Council
NSCID	National Security Council Intelligence Directive
ODM	Office of Defense Mobilization
ORR	Office of Research and Reports
OSA	Office of Special Activities
OSI	Office of Scientific Intelligence
PBCFIA	President's Board of Consultants on Foreign Intelligence Activities
P-E	Perkin-Elmer Company
PFIAB	President's Foreign Intelligence Advisory Board
PI	Photointerpreter
PIC	Photographic Intelligence Center
PID	Photo-Intelligence Division
PSAC	President's Science Advisory Committee
RAF	Royal Air Force
RFP	Request for proposal
SAB	Scientific Advisory Board (USAF)
SAC	Science Advisory Committee
SAC	Strategic Air Command
SA/PC/DCI	Special Assistant to the DCI for Planning and Coordination
SAM	Surface-to-air missile
SEI	Scientific Engineering Institute
SENSINT	Sensitive intelligence (USAF)
SLAR	Side-looking aerial radar
TAS	True air speed
TCP	Technological Capabilities Panel
USIB	United States Intelligence Board
WADC	Wright Air Development Command (USAF)
WRSP	Weather Reconnaissance Squadron, Provisional

APPENDIX B

Key Personnel

AYER, Frederick, Jr.
Special assistant to Trevor Gardner in the Office of the Secretary of the Air Force, Ayer was a strong advocate of overhead reconnaissance by balloons and an early supporter of Lockheed's CL-282 design.

BAKER, James G.
Harvard astronomer and lens designer, Baker was a leading designer of high-acuity aerial lenses during World War II and continued this work after the war. He also headed the Air Force Intelligence Systems Panel and served on the Technological Capabilities Panel's Project Three committee that urged the development of the U-2 aircraft. Baker designed the lenses for the U-2's cameras.

BISSELL, Richard M., Jr.
Head of all CIA overhead reconnaissance programs from 1954 until 1962, a former economics professor at MIT and high official of the Marshall Plan, Bissell became Allen W. Dulles's Special Assistant for Planning and Coordination in January 1954 and received responsibility for the new U-2 project at the end of that year. Later he also headed the first photosatellite project and oversaw the development of the OXCART. In 1959 Bissell became Deputy Director for Plans but kept the reconnaissance projects under his control. He resigned from the CIA in February 1962.

CABELL, George Pearre
Air Force general and DDCI from 1953 until 1962. Because of Cabell's many years of experience in aerial reconnaissance, DCI Dulles delegated most of the responsibility for the reconnaissance projects to him.

CARTER, Marshall S.
Army general who served as DDCI from 1962 until 1965. During the period leading up to the Cuban Missile Crisis, Carter served as Acting DCI on a number of occasions while DCI McCone was out of town. In October 1962 he fought unsuccessfully to keep the CIA involved in flying reconnaissance missions over Cuba. Carter became the Director of the National Security Agency in 1965.

CHARYK, Joseph R.
An aeronautical engineer who had followed careers first in academia and then the aerospace industry, Charyk became the Chief Scientist of

the Air Force in January 1959. Five months later he moved up to Assistant Secretary of the Air Force for Research and Development, and the following year he became Under Secretary of the Air Force. In these positions he was involved in coordination with the CIA on both the U-2 and OXCART projects. In 1963 Charyk left-government to become the first chairman of the Communications Satellite Corporation.

CUNNINGHAM, James A., Jr.
An ex–Marine Corps pilot, he became the administrative officer for the U-2 project in April 1955. Cunningham handled the day-to-day management of the U-2 program and brought only the more complex problems to Richard Bissell's attention. Later he served as the Deputy Director of the Office of Special Activities and then Special Assistant to the Deputy Director for Science and Technology.

DONOVAN, Allen F.
An aeronautical engineer who had helped to design the P-40 fighter while working at the Curtiss-Wright Corporation, Donovan was one of the founders of the Cornell Aeronautical Laboratory after World War II. He served on several Air Force advisory panels and was a strong advocate of the proposed Lockheed CL-282 aircraft. Later he became vice president of the Aerospace Corporation.

DOOLITTLE, James H.
A vice president of Shell Oil Company and an Army Air Force reserve general, Doolittle headed General Eisenhower's Air Staff during World War II. After the war Doolittle served on many Air Force advisory panels, and in 1954 he chaired a special panel investigating the CIA's covert activities. Doolittle also served on the Technological Capabilities Panel and the President's Board of Consultants on Foreign Intelligence Activities.

DUCKETT, Carl E.
Headed the Directorate of Science and Technology from September 1966 until May 1976, first as Acting Deputy Director and then as Deputy Director beginning in April 1967. During his tenure, the emphasis in the CIA's overhead reconnaissance program shifted from aircraft to satellites.

DULLES, Allen W.
DCI from 1953 until 1961. Although initially reluctant to see the CIA involved in aerial reconnaissance, which he viewed as the military's area of responsibility, Dulles became a strong supporter of the U-2

program when he saw how much intelligence it could gather on the Soviet Union. Because his own interests lay more in the area of human intelligence, he left the management of the reconnaissance program in the hands of DDCI Cabell and project director Richard Bissell.

GARDNER, Trevor
During World War II, Gardner worked on the Manhattan Project, and later he headed the General Tire and Rubber Company before starting his own research and development firm, the Hycon Company, which built aerial cameras. Gardner served as the Secretary of the Air Force's Special Assistant for Research and Development and then as the Assistant Secretary for Research and Development during Eisenhower's first term of office. Gardner's concern about the danger of a surprise attack helped lead to the establishment of the Technological Capabilities Panel. Gardner also urged the building of Lockheed's CL-282 aircraft.

GEARY, Leo P.
Air Force colonel (later brigadier general) who was James Cunningham's Air Force counterpart in the U-2 program. He was instrumental in diverting engines from other Air Force projects for use in the U-2, and his 10 years with the U-2 project provided a high degree of continuity.

GOODPASTER, Andrew J.
An Army colonel who served as President Eisenhower's Staff Secretary from 1954 to 1961. During this period, he was the CIA's point of contact in the White House for arranging meetings with the President on the subject of overhead reconnaissance. Goodpaster's later career included service as the supreme commander of NATO and then commandant of the US Military Academy at West Point.

HELMS, Richard M.
DCI from 1966 to 1973. During his tenure as DCI, the CIA's manned reconnaissance program came under heavy pressure because of competition from the Air Force's reconnaissance program.

JOHNSON, Clarence L. (Kelly)
One of the nation's foremost aeronautical designers, Kelly Johnson graduated from the University of Michigan's School of Aeronautics in 1933 and began working for the Lockheed Aircraft Corporation. During World War II he designed the P-38 fighter, and after the war his design successes continued with the F-104 jet fighter, the Constellation airliner, and the CIA's two strategic reconnaissance aircraft, the U-2 and the OXCART A-12.

KIEFER, Eugene P.
An Air Force officer with a degree in aeronautical engineering who in 1953 informed a friend at Lockheed of the Air Force's search for a high-altitude reconnaissance aircraft, thus, leading to the initial design of the CL-282. After leaving the Air Force, Kiefer became Richard Bissell's technical adviser for the OXCART and photosatellite programs.

KILLIAN, James R., Jr.
President of the Massachusetts Institute of Technology, Killian headed a high-level and very secret study of the nation's ability to withstand a surprise attack. While this project was still under way, he and Edwin Land persuaded President Eisenhower to support the development of a high-altitude reconnaissance aircraft, the U-2. Later, Killian headed Eisenhower's Board of Consultants for Foreign Intelligence Activities, served as his Cabinet-level science adviser, and chaired the President's Science Advisory Board. Killian was also chairman of the President's Foreign Intelligence Advisory Board under John F. Kennedy.

LAND, Edwin H.
An extremely talented inventor famous for the development of polarizing filters and the instant-film camera. Land also devoted considerable time and energy to voluntary government service. During World War II, Land worked for the Radiation Laboratories, and after the war he served on numerous Air Force advisory panels. As the head of the Technological Capabilitites Panel's study group investigating US intelligence-gathering capabilities, Land became a strong advocate of the development of a high-altitude reconnaissance aircraft (the CL-282) under civilian rather than Air Force control. Land and James Killian persuaded President Eisenhower to approve the U-2 project and later the first photosatellite project. Land also served on the President's Board of Consultants for Foreign Intelligence Activities.

LEGHORN, Richard S.
An MIT graduate in physics, Leghorn joined the Army Air Force in 1942 and went to work for reconnaissance expert Col. George Goddard. By the time of the invasion of Europe, Leghorn was chief of reconnaissance for the 9th Tactical Air Force. After the war, Leghorn began preaching the need for "pre-D-day" reconnaissance in order to gather intelligence on the Soviet Bloc. He returned to the Air Force during the Korean war and later worked for Harold Stassen's Disarmament Office. In 1956 he became the head of the Scientific

Engineering Institute, an Agency proprietary working on ways to reduce the U-2's vulnerability to radar detection. In 1957 he founded Itek Corporation.

LUNDAHL, Arthur E.
A Navy photointerpreter during World War II and afterward. Lundahl became the chief of the Photo-Intelligence Division in 1953. To support the U-2 project, he established a separate photointerpretation center under Project HTAUTOMAT. Under his leadership the Photo-Intelligence Division grew rapidly and achieved office status as the Photographic Intelligence Center in 1958. In 1961 Lundahl became the first head of the National Photograhic Interpretation Center, which combined the photointerpretation efforts of the CIA and the military services.

McCONE, John A.
DCI from 1961 to 1965. A strong supporter of the CIA's manned reconnaissance program, McCone presided over the OXCART's main period of development and pushed for a greater role for the CIA in its joint reconnaissance programs with the Department of Defense.

MILLER, Herbert I.
Miller worked in the Office of Scientific Intelligence's nuclear branch and became Richard Bissell's first deputy for the U-2 project. He later left the Agency to work for the Scientific Engineering Institute.

NORTON, Garrison
An assistant to Trevor Gardner, Norton became an early supporter of the Lockheed CL-282 and started the CIA's interest in overhead reconnaissance by informing Philip Strong about the aircraft. Norton later became Navy Assistant Secretary for Research and Development and was involved with the OXCART program.

OVERHAGE, Carl F. J.
After working on the development of Technicolor, Overhage went to work for Kodak. He headed the Beacon Hill Panel in 1952 and later became director of Lincoln Laboratories.

PARANGOSKY, John N.
Parangosky worked for Richard Bissell's Development Projects Staff in the mid-1950s. He served as deputy chief of the Adana U-2 unit in 1959 and became project manager of the OXCART program from its inception through the test flight stage.

PERKIN, Richard S.
President of the Perkin-Elmer Corporation, Perkin was a close friend of James Baker and was also a member of several advisory panels, including the BEACON HILL project. He helped Baker decide what cameras to use in the first U-2 aircraft.

POWERS, Francis Gary
An Air Force Reserve Officer who became a CIA U-2 pilot in 1956, Powers flew 27 successful missions before being shot down over the Soviet Union on 1 May 1960. After his return to the United States in exchange for Soviet spymaster Rudolf Abel in 1962, Powers was cleared of all allegations of misconduct in his mission, capture, trial, and captivity. He became a test pilot for Lockheed and later piloted light aircraft and helicopters for radio and television stations. He died in a helicopter crash on 1 August 1977.

PURCELL, Edward M.
A physicist who won a Nobel prize in 1954 for his work in nuclear resonance, Purcell served on a number of advisory bodies, including the USAF Scientific Advisory Committee and Edwin Land's Technological Capabilities Panel study group. It was Purcell's ideas for reducing the radar cross section of the U-2 that led to the OXCART program. Purcell also contributed to the satellite programs.

RABORN, William F., Jr.
DCI from 1965 to 1966, Raborn pushed for the deployment of OXCART to the Far East but failed to sway the top officials of the Johnson administration.

REBER, James Q.
After serving as the Assistant Director for Intelligence Coordination in the early 1950s, Reber became the chairman of the Ad Hoc Requirements Committee in 1955 and continued to chair this committee after it was taken over by the US Intelligence Board in 1960 and renamed the Committee on Overhead Requirements. In 1969 he became the chairman of the USIB's SIGINT Committee.

RODGERS, Franklin A.
Formerly of MIT, Rodgers was the chief engineer at the Scientific Engineering Institute who converted the theories of Edward Purcell into practical systems to reduce the radar image of the U-2 and especially the OXCART.

SCHLESINGER, James R.
DCI from February to July 1973, Schlesinger supported the Nixon administration's proposal to terminate the Agency's U-2 program.

SCOTT, Roderic M.
An engineer with Perkin-Elmer who worked with James Baker in designing the first cameras for use in the U-2. Scott helped design the 30001 camera for the OXCART.

SCOVILLE, Herbert, Jr.
In February 1962 Scoville became the first Deputy Director for Research, which took over control of the Agency's reconnaissance programs from the Deputy Director for Plans. Frustrated by the lack of support from the DCI and the other directorates, he resigned in June 1963.

SEABERG, John
An aeronautical engineer who was recalled to active duty with the Air Force during the Korean war, Seaberg drafted the first specifications for a high-flying jet reconnaissance aircraft in 1953.

STEVER, H. Guyford
A professor of aeronautical engineering at MIT, Stever served on numerous Air Force advisory panels and later became the Air Force's chief scientist.

STRONG, Philip G.
Chief of collection in the Office of Scientific Intelligence, Strong kept himself well informed on developments in overhead reconnaissance and attended many Air Force advisory panel meetings as an observer. In 1954 he learned about the Lockheed CL-282 design and passed the information on to Edwin Land's study group investigating US intelligence-gathering capabilities.

WHEELON, Albert ("Bud") D.
Wheelon became the Deputy Director for Science and Technology in August 1963 following the reorganization and renaming of the Deputy Director for Research. He held this position until September 1966.

APPENDIX C

Electronic Devices Carried by the U-2

From the beginning the U-2 was envisioned as more than a camera platform. In fact, the U-2 would ultimately carry only five types of photographic equipment but more than 20 different types of electronic devices, some for collecting electronic intelligence (ELINT), others ("ferret" equipment) for gathering intelligence on foreign radars, and a few for self-protection—electronic countermeasures to defeat enemy missile-control radars.

The various electronic, countermeasures and intelligence-gathering systems designed for the U-2 received designations using Roman numerals—Systems I through XXII. The first seven devices were built by the Ramo-Wooldridge firm, now part of the TRW Corporation. System-I used S- and X-band ELINT receivers to collect ground-controlled intercept and air defense signals. Weighing only 7.7 kilograms, this system was aboard all U-2s from 1955 through 1959. System-II, a communications and navigation system, never worked properly and was canceled. System-III, a 16-kilogram VHF recorder for communications intelligence (COMINT), was never used and was transferred to the Navy in 1958. System-IV, a ferret device that recorded electromagnetic energy in the 150- to 4,000-MHz range was used on 16 missions between 1957 and 1959, when it was given to the Air Force. System-V was similar to System-I but covered nine wave bands. The device was so heavy that U-2s using it could not carry a camera system. System-V was used on only three missions and was replaced by the lighter weight System-VI that covered the P-, L-, S-, and X-band frequencies and could be used with either the A or B camera. System-VI was used from 1959 through 1966.

The growing need for data on Soviet missile development led to a contract with the firm of Haller-Raymond-Brown (HRB) to build a missile-telemetry intercept system as quickly as possible. The resulting device (System-VII) could record up to 12 minutes of data from six simultaneous frequencies. This unit first saw service on 9 June 1959 and was used on another 22 missions during the next year. Following modifications to make System-VII suitable for use by the Navy, the designation was changed to System-VIII.

System-IX was an electronic-countermeasures (ECM) device for generating false-angle information in response to X-band radar pulses from surface-to-air missile radars. Also known as the Mark-30, the

unit was manufactured by the Granger Company. One of these devices was aboard Gary Powers' U-2 when he was shot down.

System-X was a modification of the HRB's System-VII that was specially built in 1962 for a mission over the Soviet Union that never took place. Systems-XI through XV were ECM devices used by U-2s overflying China and North Vietnam during the Vietnam war. System-XVI was a passive ELINT collector.

System-XVII was built by HRB-Singer as a result of an October 1963 USIB requirement for the collection of antiballistic-missile (ABM) data from Saryshagan. The system was to be deployed in a U-2 that would fly over western China, along the Sino-Soviet border, collecting data on the ABMs being tested at Saryshagan. By the time the unit was completed in 1965, however, the tipoff time before test launches had been reduced from almost 24 hours to less than an hour, making it impossible to stage U-2 missions in time to collect the data.

In the late 1960s, additional ECM systems were needed to counter the increasing threats posed by more accurate SAMs and higher flying aircraft. System-XX was specifically designed to counter the acquisition and guidance radars used by MIG aircraft, and System-XXII was an infrared jammer to counter air-to-air missiles. System-XXI, a COMINT package that replaced the much older System-III, was originally developed for the OXCART program and was later adapted for use aboard the follow-on U-2, the U-2R.

APPENDIX D

U-2 Overflights of the Soviet Union, 4 July 1954–1 May 1960

Date	Mission	Pilot	Airfield	Unit	Payload	Route
4 July 1956	2013	Stockman	Wiesbaden	A	A-2	East Germany, Poland, Minsk, Leningrad, Estonia, Latvia, Poland
5 July 1956	2014	Vito	Wiesbaden	A	A-2	East Germany, Warsaw, Minsk, Moscow, Estonia, Latvia, Lithuania, Poland
9 July 1956	2020	Knutson	Wiesbaden	A	A-2	East Germany, Poland, Minsk, Poland
9 July 1956	2021	Overstreet	Wiesbaden	A	A-2	Czechoslovakia, Vienna, Hungary, L'vov, Kiev, Minsk, Poland
10 July 1956	2024	Dunaway	Wiesbaden	A	A-2	Poland, Kishinev, Kerch', Sevastopol', Simferopol', Odessa, Romania, Hungary
20 November 1956	4016	Powers	Adana	B	A-2	Iran, Yerevan, Baku, Astara, Caucasus
18 March 1957	4020	Cherbonneaux	Adana	B	Sys-V	Soviet border to Afghanistan, Azerbaijan, Armenia, Georgia
20 June 1957	6005	Rand	Eielson	C	B	Khaylyulya, Ust'-Kamchatsk, Kozyrevsk, Karaganskiy-Ostrov
5 August 1957	4035	Edens	Lahore	B	B	Afghanistan, Tashkent, Tyuratam, Kazalinsk, Aral Sea
11 August 1957	4039	McMurray	Lahore	B	B	Alma-Ata, Ust'-Kamenogorsk, Sinkiang
21 August 1957	4045	Snider	Lahore	B	A-2	Novokuznetsk, Tomsk
21 August 1957	4048	Jones	Lahore	B	A-2	Lake Balkhash, Karaganda, Omsk, Alma-Ata

APPENDIX D

U-2 Overflights of the Soviet Union, 4 July 1954–1 May 1960 (continued)

Date	Mission	Pilot	Airfield	Unit	Payload	Route
22 August 1957	4049	Birkhead	Lahore	B	A-2	Merket Bazar, Kuldja, Abakan, Krasnoyarsk, Kansk, Sinkiang
22 August 1957	4050	Cherbonneaux	Lahore	B	A-2	Lake Balkhash, Semipalatinsk, Barnaul, Prokop'yevsk, Novokuznetsk, Leninogorsk
28 August 1957	4058	Jones	Lahore	B	A-2	Dushanbe, Tashkent, Tyuratam, Kazalinsk, Aral Sea
10 September 1957	4059	Hall	Adana	B	A-2	Krasnovodsk, Gur'yev, Astrakhan', Tbilisi
16 September 1957	6008	Baker	Eielson	C	A-2	Kamchatka Peninsula, Milkovo
13 October 1957	2040	Stockman	Giebelstadt	A	A-2	Norway, Finland, Murmansk, Kandalaksha
1 March 1958	6011	Crull	Atsugi	C	A-2	Dal'nerechensk, Khabarovsk, Blagoveshchensk, Belagorsk, Komsomolsk, Sovetskaya Gavan'
9 July 1959	4125	Knutson	Peshawar	B	B	Tyuratam for suspected Sputnik launch
6 December 1959	8005	Robinson [a]	Peshawar	B	B	Kuybyshev, Kapustin Yar, Caucasus
5 February 1960	8009	MacArthur [a]	Peshawar	B	B	Tyuratam, Kazan
9 April 1960	4155	Erickson	Peshawar	B	B	Lake Balkhash, Semipalatinsk, Kyzylespe, Dzhezhkazgan, Tyuratam
1 May 1960	4154	Powers	Peshawar	B	B	Tyuratam, Kyshtym, Sverdlovsk; downed by SAM

[a] RAF pilot

APPENDIX E

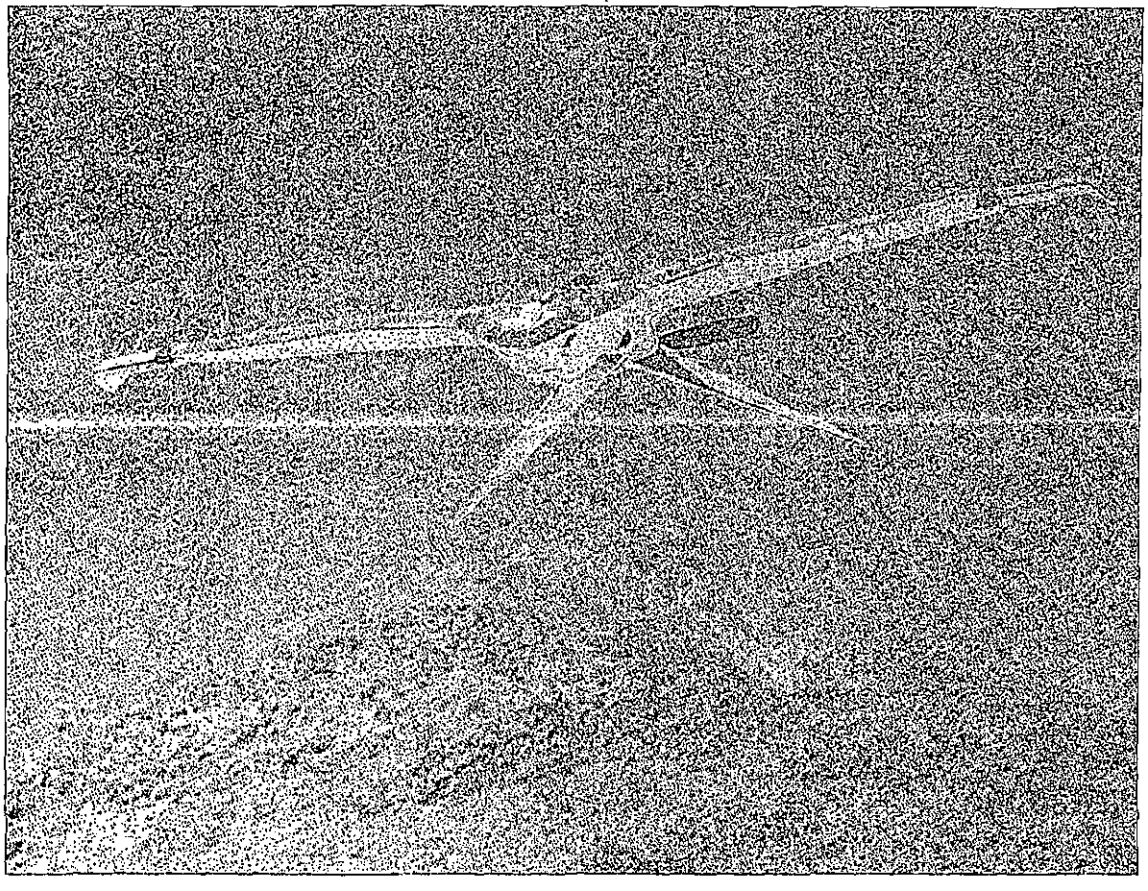
Unmanned Reconnaissance Projects

AQUILINE

In the early 1960s, there were many problems in obtaining coverage of hostile territory. The U-2 was too vulnerable to Soviet surface-to-air missiles, as had been demonstrated by losses over the Soviet Union, Cuba, and the People's Republic of China. The OXCART was still under development and even when completed might prove vulnerable to Soviet radars and missiles. Although safe from interception, the newly developed photosatellites could not provide coverage of a desired target on short notice. Because several of the intelligence community's primary targets such as Cuba and the new Soviet radar installation at Tallinn (Estonia) were not located deep in hostile territory, CIA scientists and engineers began to consider the possibility of using small, unmanned aircraft for aerial reconnaissance. They believed that recent advances in the miniaturization of electronic technology would make possible the development of a reconnaissance vehicle with a very-low-radar cross section and small visual and acoustical signatures. Such a vehicle could reconnoiter an area of interest without the hostile country realizing that it had been overflown.

In mid-1965, David L. Christ, chief of the Office of Research and Development's Applied Physics Division, and Frank Briglia of the same office began working on the concept of a small, inexpensive aircraft that would be about the size of a large bird and could carry various payloads for photography, nuclear sensing, and ELINT collection. ORD soon formed a Special Projects Group to develop the aircraft with Briglia as the project manager. Only one firm—the Douglas Aircraft Company—responded favorably to a request for proposal to study the feasibility of a low-altitude reconnaissance system. On 15 November 1965, Douglas received a study contract. This was followed by two Agency contracts on 21 November 1966 for the development of an operational low-altitude intelligence-gathering system. Further contracts followed in 1968 and 1969.

The AQUILINE prototype developed by Douglas Aircraft (which became part of McDonnell-Douglas in 1969 as the result of a merger) was essentially a powered glider with an 8.5-foot wingspan. The aircraft weighed only 105 pounds. AQUILINE's tail-mounted engine drove a two-bladed propeller. Powered by a small 3.5-horsepower two-cycle engine originally developed by the McCullough



Project AQUILINE

Corporation for chainsaws, the aircraft's designed speed was 60 knots at an altitude of 1,000 feet with a 15-pound payload. The aircraft could fly at this speed for up to 30 hours, thanks to the engine's extremely high fuel efficiency: 480 nautical miles per gallon of fuel.

In 1968 tests on an AQUILINE prototype at Randsburg Wash on the US Navy's Naval Ordnance Test Station at China Lake, California, showed that the aircraft was extremely difficult to see. To assist pilots of chase aircraft in keeping AQUILINE in sight, its entire upper surface was painted bright orange; even so, sighting remained difficult.

The testing process was very hard on AQUILINE because it was recovered by flying it into a net close to the ground, which almost always caused some damage to the wings or propeller. As a result, one

or more of the aircraft was always being repaired, and eventually three of the five AQUILINE prototypes were destroyed in testing.

Although AQUILINE's visual and accoustic signatures were very small, its radar cross section continued to cause problems. The radar cross section at low frequencies was less than one-tenth of a square meter, giving it a radar signature smaller than that of an eagle, but there were flares in the VHF region that increased the likelihood of detection.

Studies indicated that the radar cross section problems could eventually be reduced to acceptable levels, but the greatest weakness of the AQUILINE project was its navigational system. AQUILINE did not have a programmable autopilot; it had to be flown by remote control from the ground. Once the aircraft flew over the horizon, all navigational commands had to be relayed to it by high-flying aircraft. A DC-6 loitering at 25,000 feet would give AQUILINE a range of 250 nautical miles; use of a U-2 at 70,000 feet could increase this to 350 nautical miles.

Such a range was suitable for the targets originally conceived for AQUILINE,

25X1, E.O.13526

and the development time and costs for AQUILINE began to soar. By 1971 ORD had spent six years and $33 million on AQUILINE but still had many important problems to resolve.

At this point the project was turned over to the Office of Special Activities for operational testing at Area 51 in Nevada. Flight tests showed the aircraft to be successful by the standards originally set for the project in 1967, as it flew 130 miles and obtained very high resolution photography of a target before returning successfully to the original launchsite. However, improving AQUILINE sufficiently to make it a practical long-range reconnaissance system was estimated to cost another $35 million and take two to three years. On the recommendation of DDS&T Carl Duckett on 1 November 1971, Project AQUILINE was canceled. The project's aircraft and equipment were eventually transferred to the US Army for use in a battlefield management system known as AQUILA, which has yet to be deployed because of continuing developmental problems.

Project AXILLARY

AXILLARY

While Project AQUILINE was still under development, its chief aerodynamicist, Charles N. Adkins, left the program because he believed that its escalating costs would prevent it from ever producing a deployable aircraft. He wanted to build a small, inexpensive remote-controlled aircraft to test a low-cost lightweight autopilot currently being developed by ORD. Under a $5,000 time-and-materials contract with Melpar, Incorporated, Adkins hired a local model aircraft builder to assemble and modify a standard Hawk-750 glider kit and power it with a rear-mounted engine and pusher propeller. Following a series of successful test flights, Adkins installed a small camera and took a number of aerial photographs.

By this time the effort to build a "Miniature Multi-Purpose Airborne Vehicle" had become known as Project AXILLARY. Melpar, Inc., received a second contract for $50,000 to install ORD's autopilot in the aircraft, and the project managers now began searching for a use for their vehicle. The two main possibilites were (1) as a short-range reconnaissance vehicle for use in a peace-monitoring or intelligence-gathering system and (2) as a short-range warhead delivery system. In 1971 the Office of Special Activities evaluated AXILLARY flight-testing and determined that the small model aircraft was not suitable for use as a covert reconnaissance vehicle because of its large radar cross section and significant accoustical signature. The aircraft's radar signature made it potentially useful as a weapons systems, however. ORD suggested that AXILLARY be equipped with a radar-homing unit and a small warhead, which would make it an inexpensive means for destroying the radars employed by surface-to-air missile systems in North Vietnam. The Director of

Defense Research and Engineering, John Foster, liked the concept and provided DOD funding for ORD to develop two versions of AXILLARY, one with a radar-homer and one with a television reconnaissance package. The radar homing system proved successful as AXILLARY sought out and destroyed a radar during testing at China Lake Naval Air Station. However, the end of US involvement in Vietnam in early 1973 led to the cancellation of further DOD funding, and AXILLARY was placed on the shelf. Melpar, Inc., continued to work on the radar-homing version and eventually sold some AXILLARY-type aircraft to Israel, which used them to destroy Syrian radars in Lebanon in 1982.

PINE RIDGE

While work was still in progress on low-altitude, short-range reconnaissance systems like AQUILINE and AXILLARY, CIA scientists and engineers were also working on a high-altitude recoverable unmanned reconnaissance vehicle with an extremely-low-radar cross section that would enable it to fly undetected over hostile territory. During the 1960s there had been a study conducted by the Teledyne Ryan Aeronautical Company of San Diego on the construction of an undetectable drone known as SANDY HOOK. This 44-foot long drone would operate at 120,000 feet at Mach 0.9, with a range of 5,000 nautical miles. Its radar cross section 07zwould be 0.01 square meters. Project SANDY HOOK never advanced very far in development; projected high costs and substantial technical risks led to its cancellation in December 1969.

Some of the concepts in SANDY HOOK were carried over into a new project known as PINE RIDGE, which was a proposal for an unmanned reconnaissance vehicle with an even lower radar cross section (0.001 square meter). Research on SANDY HOOK had indicated that a radar cross section this low was attainable and would prevent detection and tracking by existing radar defense systems. The PINE RIDGE proposal called for a delta-shaped vehicle, approximately 17 feet long with a 21-foot wingspan. Two vertical stabilizers would give the vehicle an overall height of three to four feet. An existing Teledyne J-100-CA-100 Turbojet engine could have been used to power the vehicle at Mach 0.9 at the operating altitude of 65,000 to 75,000 feet. Range was estimated at 3,300 nautical miles.

Despite interest within the CIA and the Air Force for an undetectable reconnaissance vehicle, PINE RIDGE was never funded. In January 1971 high-level representatives from DOD and CIA rejected a proposal for a feasibility study to be conducted by Ryan Aircraft.

The following year, the Director of the Office of Special Activities, Brig. Gen. Harold F. Knowles, wrote a memorandum to the DCI proposing that the CIA develop a clandestine low-radar-cross-section vehicle like PINE RIDGE, but this proposal also failed to attract support. With all the improvements in satellite reconnaissance that had taken place during the previous decade, neither CIA nor the Air Force was willing to invest substantial funding in a project to send aircraft, manned or unmanned, over the Soviet Union.

BIBLIOGRAPHY

Published Works Cited

Aart, Dick van der. *Aerial Espionage, Secret Intelligence Flights by East and West*. Shrewsbury, England: Airlife Publishing, 1985.

"AF Cites Red Bomber Progress." *Aviation Week*, 24 May 1954, p. 14.

"Alleged Overflight of Soviet Area by American Planes." *US Department of State Bulletin*, vol. 36, 28 January 1957, p. 135.

"Alleged Violations of Soviet Territory: Soviet Note of July 10, 1956 with U.S. Reply." *US Department of State Bulletin*, 30 July 1956, pp. 191-192.

Alsop, Stewart. *The Center*. New York: Popular Library, 1968.

Ambrose, Stephen E. *Eisenhower*. Vol. 2, *The President*. New York: Simon and Schuster, 1984.

Andrew, Christopher. *Her Majesty's Secret Service: The Making of the British Intelligence Community*. New York: Viking, 1986.

"Aviation Week Story Spurs Debate on US, Red Airpower Positions." *Aviation Week*, 30 May 1955, pp. 13-14.

Beschloss, Michael R. *Mayday: Eisenhower, Khrushchev and the U-2 Affair*. New York: Harper & Row, 1986.

Brown, William H. "J58/SR-71 Propulsion Integration." *Studies in Intelligence* 26 (Summer 1982): pp. 17-18 (Unclassified).

Brugioni, Dino. *The Cuban Missile Crisis-Phase I, 29 August-16 October 1962*. DDS&T Historical Series, NPIC-1 (CIA: NPIC, 1971) ~~(Secret)~~.

Burrows, William E. *Deep Black: Space Espionage and National Security*. New York: Random House, 1987.

"Can Soviets Take the Air Lead? What LeMay, Wilson, Ike Say." *US News and World Report*, 11 May 1956, pp. 108-114.

Chronology of the Office of Special Activities, 1954-1968. (CIA: DS&T, 1969) ~~(Top Secret~~ Codeword).

Coughlin, William. "Gardner Defends Greater R&D Spending." *Aviation Week*, 26 September 1955, p. 14.

Eastman, Ford. "Defense Officials Concede Missile Lag." *Aviation Week*, 9 February 1959, pp. 26-27.

Edwards, Philip K. "The President's Board: 1956-60." *Studies in Intelligence* 13 (Summer 1969): p. 118 ~~(Secret)~~.

Eisenhower, Dwight D. *Waging Peace, 1956-1961*. Garden City, NY: Doubleday, 1965.

Freedman, Lawrence. *US Intelligence and the Soviet Strategic Threat*. 2nd ed. Princeton: Princeton University Press, 1986.

George, A. L. *Case Studies of Actual and Alleged Overflights, 1930-1953*. Rand Study RM-1349. Santa Monica: Rand, 1955 ~~(Secret)~~.

[redacted] *The Defector Reception Center Germany, 1951 to 1967*. Clandestine Service Historical Series CSHP-41. CIA: History Staff, 1972 ~~(Secret)~~.

Hotz, Robert. "Firing of 900-Mile Russian Missile Spurs US Changes." *Aviation Week*, 20 February 1956, pp. 26-28.

———. "Russian Air Force Now Gaining in Quality." *Aviation Week*, 12 March 1956, pp. 286-291.

———. "Russian Jet Airpower Gains Fast on US." *Aviation Week*, 23 May 1955, pp. 12-15.

"Is Russia Ahead in Missile Race?" *US News and World Report*, 6 September 1957, pp. 30-33.

"Is Russia Really Ahead in Missile Race?" *US News and World Report*, 4 May 1956, pp. 34-35.

"Is Russia Winning the Arms Race?" *US News and World Report*, 18 June 1954, pp. 28-29.

"Is US Really Losing in the Air?" *US News and World Report*, 18 May 1956, pp. 25-27.

Withheld under statutory authority of the Central Intelligence Agency Act of 1949 (50 U.S.C., section 403g)

Johnson, Clarence L. "Development of the Lockheed SR-71 Blackbird." *Studies in Intelligence* 26 (Summer 1982): p. 4 (Unclassified).

Johnson, Clarence L. with Maggie Smith. *Kelly: More Than My Share of It All.* Washington, DC: Smithsonian Institute Press, 1985.

Khrushchev, Nikita S. *Khrushchev Remembers: The Last Testament.* Boston: Little, Brown, & Co., 1974.

Killian, James R., Jr. *Sputnik, Scientists, and Eisenhower: A Memoir of the First Special Assistant to the President for Science and Technology.* Cambridge: MIT Press, 1977.

Kistiakowsky, George B. *A Scientist at the White House.* Cambridge: Harvard University Press, 1976.

Kleyla, Helen H., and Robert D. O'Hern. *History of the Office of Special Activities, DS&T.* 16 vols. Directorate of Science and Technology Historical Series OSA-1. CIA: DS&T, 1969 ~~(Top Secret~~ Codeword).

Lay, James S. "The United States Intelligence Board, 1958-1965." CIA History Staff MS-2 (draft), 1974 ~~(Top Secret~~ Codeword).

Licklider, Roy E. "The Missile Gap Controversy." *Political Science Quarterly* 85 (1970): pp. 600-615.

Linehan, Urban J. *The National Photographic Interpretation Center.* Vol. 1, *Antecedents and Early Years.* Directorate of Science and Technology Historical Series NPIC-2. CIA: NPIC, 1972 ~~(Secret).~~

———. *National Photographic Interpretation Center: The Years of Project HTAUTOMAT, 1956-1958.* 6 vols. Directorate of Science and Technology Historical Series NPIC-3. CIA: NPIC, 1974 ~~(Secret~~).

Miller, Jay. *Lockheed U-2.* Aerograph 3. Austin, Texas: Aerofax, 1983.

Minnich, Thomas P. "The OXCART Story." *Studies in Intelligence* 15 (Winter 1971): pp. 1-34 ~~(Secret~~).

"Missiles Away." *Time*, 30 January 1956, pp. 52-55.

"Mystery of the RB-47." *Newsweek*, 25 July 1960, pp. 36-37.

"Nikita and the RB-47." *Time*, 25 July 1960, pp. 30-31.

"Office of Special Activities History, April 1969 to Phase-Out." (draft) (CIA: DS&T, 1974), chap. 3, pp. 36-42 ~~(Top Secret~~ Codeword).

Pero, Richard K. "Message Received—Unfortunately." *Studies in Intelligence* 27 (Winter 1983): p. 29 ~~(Secret).~~

Pocock, Chris. *Dragon Lady: The History of the U-2 Spyplane*. Shrewsbury, England: Airlife, 1989.

Powers, Francis Gary, with Curt Gentry. *Operation Overflight*. New York: Holt, Rinehart and Wilson, 1970.

Powers, Thomas. *The Man Who Kept the Secrets: Richard Helms and the CIA*. New York: Alfred A. Knopf, 1979.

Prados, John. *Presidents' Secret Wars: CIA and Pentagon Covert Operations Since World War II*. New York: William Morrow, 1986.

———. *The Soviet Estimate: U.S. Intelligence Analysis and Russian Military Strength*. New York: Dial Press, 1982.

[redacted] *Covert Support to Indonesian Revolutionary Government, 1957-1958*. 2 vols. Clandestine Service Historical Series, CSHP-53. CIA: History Staff, 1970 ~~(Secret).~~

"Red Air Force: The World's Biggest." *Newsweek*, 23 August 1954, pp. 28-33.

Richelson, Jeffrey. *American Espionage and the Soviet Target*. New York: William Morrow, 1987.

"Russia Parades Airpower as 'Big Stick'." *Aviation Week*, 28 June 1954, p. 15.

[redacted] *Low-Level Technical Reconnaissance over Mainland China (1955-66)*. Clandestine Services Historical Program, CSHP-2.348. CIA: History Staff, 1972 ~~(Secret).~~

Stanley, Roy M., II. *World War II Photo Intelligence*. New York: Scribners, 1981.

Withheld under statutory authority of the Central Intelligence Agency Act of 1949 (50 U.S.C., section 403g)

Sturm, Thomas A. *The USAF Scientific Advisory Board: Its First Twenty Years, 1944-1964*. Washington, DC: USAF Historical Office, 1967.

Twining, Nathan F. *Neither Liberty nor Safety*. New York: Holt, Rinehart & Winston, 1966.

United States Air Force. Project LINCOLN. *BEACON HILL Report: Problems of Air Force Intelligence and Reconnaissance*. Massachusetts Institute of Technology, 15 June 1952 ~~(Secret,~~ downgraded to Confidential).

United States Congress. Senate. Foreign Relations Committee. *Executive Sessions of the Senate Foreign Relations Committee (Historical Series)*, vol. 12, 86th Congress, 2nd Session, "Report on the U-2 Incident," 6 March 1962 (declassified 1982).

United States Congress. Senate. Select Committee to Study Governmental Operations with Respect to Intelligence Activities. *Foreign and Military Intelligence*. Book 1. Washington, DC: US Government Printing Office, 1976.

United States Office of Defense Mobilization. Science Advisory Committee. Technological Capabilities Panel. *Meeting the Threat of Surprise Attack*, 14 February 1955. ~~(Top Secret~~/Restricted Data, downgraded to ~~Secret)~~.

"US Representations to the Soviet Government on C-130 Transport Shot Down by Soviet Fighter Aircraft." *US Department of State Bulletin*, 23 February 1959, pp. 262-271.

Weber, Karl H. *The Office of Scientific Intelligence, 1949-68*. Directorate of Science and Technology Historical Series OSI-1. CIA: DS&T, 1972 ~~(Top Secret~~ Codeword).

Welzenbach, Donald E. "Observation Balloons and Weather Satellites." *Studies in Intelligence* 30 (Spring 1986): pp. 21-28 ~~(Secret)~~.

White, James J. "Francis Gary Powers—The Unmaking of a Hero, 1960-1965." (draft) CIA: History Staff, 1974 ~~(Secret)~~.

Wise, David and Thomas B. Ross. *The U-2 Affair*. New York: Random House, 1962.

Witze, Claude. "Russians Outpacing US in Air Quality, Twining Warns Congress," *Aviation Week*, 27 February 1956, pp. 26-28.

"Yakovlev Yak-25RM Mandrake," *Jane's Defence Weekly*, vol. 3, no. 7, 16 February 85.

Interviews and Speeches

Interview with Robert Amory, Jr., by Donald E. Welzenbach and Gregory W. Pedlow, Washington, DC, 22 April 1987 ~~(Secret).~~

Interview with James G. Baker by Donald E. Welzenbach, Washington, DC, 24 April 1985 ~~(Secret)~~.

Speech given by Richard Bissell at CIA Headquarters, 12 October 1965 ~~(Top Secret~~ Codeword).

Interview with Joseph V. Charyk by Donald E. Welzenbach, Washington, DC, 5 December 1984 ~~(Top Secret~~ Codeword).

Interview with former U-2 pilots James Cherbonneaux, Carmine Vito, and Hervey Stockman, by Donald E. Welzenbach, Washington, DC, May 1986 ~~(Secret).~~

Interview with James A. Cunningham, Jr., by Donald E. Welzenbach, Washington, DC, 4 October 1983 ~~(Top Secret~~ Codeword).

Interview with Allen F. Donovan by Donald E. Welzenbach, Corona del Mar, California, 20 May 1985 ~~(Secret)~~.

Interview with Brig. Gen. Leo A. Geary, USAF (Ret.), by Donald E. Welzenbach, 3 April 1986 ~~(Secret)~~.

Interview with Gen. Andrew J. Goodpaster by Donald E. Welzenbach and Gregory W. Pedlow, Washington, DC, 8 July 1987 ~~(Secret)~~.

Speech by Maj. Gen. Patrick J. Halloran (former Air Force U-2 pilot) at the National Air & Space Museum, 24 April 1986 (Unclassified).

Interview with James R. Killian, Jr. by Donald E. Welzenbach, Cambridge, Massachusetts, 2 November 1984 ~~(Secret)~~.

Interview with Edwin H. Land by Donald E. Welzenbach, Cambridge, Massachusetts, 17 and 20 September 1984 ~~(Top Secret~~ Codeword).

Interview with Brig. Gen. Jack C. Ledford, USAF (Ret.), by Gregory W. Pedlow, Washington, DC, 20 February 1987 ~~(Secret)~~.

Interview with Richard S. Leghorn by Donald E. Welzenbach, Washington, DC, 19 August 1985 ~~(Secret)~~.

Interview with Arthur S. Lundahl and Dino Brugioni by Donald E. Welzenbach, Washington, DC, 14 December 1983 ~~(Top Secret~~ Codeword) .

Interview with Garrison Norton by Donald E. Welzenbach, Washington, DC, 23 May 1983 ~~(Secret)~~.

Interview with John Parangosky by Donald E. Welzenbach, 6 March 1986 ~~(Secret)~~.

Interview with James Q. Reber by Donald E. Welzenbach and Gregory W. Pedlow, Washington, DC, 21 May 1987 ~~(Secret)~~.

Interview with Carmine Vito by Donald E. Welzenbach, Washington, DC, 7 May 1986 ~~(Secret)~~.

Interview with John S. Warner, Office of the General Counsel, by Donald E. Welzenbach, Washington, DC, 5 August 1983 ~~(Top Secret~~ Codeword).

A

A-1 camera
50
A-2 aircraft
262-263
A-2 camera
53, 77, 100, 104-105, 108
A-3 aircraft
263
A-11
268, 294
A-12
271, 273-274, 277-278, 281, 283-286, 289-292, 294-297, 299, 302-305, 307-311, 329
Abel, Rudolf
183, 332
Ad Hoc Requirements Committee
81, 114, 188-189, 332
Adana, Turkey
113
Adenauer, Konrad
101
Adkins, Charles N.
342
Administration (East) Building
42
advanced antiballistic-missile (ABM)
139
aerial cameras
13, 31, 49, 329
aerial lenses
327
aerial photography
3, 13, 33, 49, 96, 119
aerial tankers
313
aerodynamic lift
248, 277
aerodynamic structural tests
274
aerodynamicist
342
Aerospace Corporation
328
Afghanistan
126, 170, 176, 189
afterburner plume
271
ailerons
47, 69
Air Research and Development Command
5, 8
Air Staff for Intelligence
154
Air Weather Service
89, 94
air-cooled reactor
239
air-inlet system
291
air-to-air missiles
285, 336
air-traffic controllers
72-73, 289
Air/Maritime Division (AMD)
73
aircraft carrier
145, 247, 250
airfields
82, 95-96, 105, 117, 153, 216, 221, 226, 317
airframe
5, 8, 10, 24, 26, 48, 56, 71, 105, 132-133, 248, 259, 274, 279, 286, 289
airframe contractors
259
airspace
3-4, 19, 127, 134, 142, 144-147, 152-153, 182, 231, 239, 241, 303, 306, 318, 322
airspeed
68, 76, 199, 235, 295
Akrotiri, Cyprus
117, [redacted]
Alaska
133-134, 145, 316

25X1, E.O.13526

Albania
140, 157
Albuquerque, New Mexico
74, 78-79
Allen, Edward L.
21
Almaza Airbase, Egypt
119
altimeters
59-60
altitude
1, 4-6, 7, 8-11, 12, 13-19, 22-26, 33-35, 39, 46-47, 49, 52, 54-55, 59, 61-64, 71-72, 74, 76, 79-80, 84, 87, 89, 93-94, 97, 101, 108, 125, 132, 142, 145, 148-149, 152, 156-157, 168, 175-179, 185, 207, 217-219, 229, 235, 240-241, 246-247, 251, 254, 259-260, 262, 263, 267, 268, 273, 277, 278-279, 284, 286, 288, 292, 294, 296-297, 304, 312-313, 320, 330, 339, 340, 343
aluminum
271
Ambon Island
214
American Civil War
1
Amory, Robert
114
Anderson, Rudolph
210
Andoya, Norway
172
"Angel",
45, 50, 66, 262
Anglo-French bombing campaign
117
Anglo-Soviet relations
94
Ankara, Turkey
135, 178
antiaircraft missile
168
antiaircraft weapons
93, 148
antiballistic missile system
298, 303
antiradar studies
267, 274
Apollo VII
254
Applied Physics Division
339
AQUATONE
40-43, 59, 62, 66, 74, 77, 82, 88, 93, 96, 100, 108, 113, 115, 122, 134, 155
AQUILA
341
AQUILINE
304, 339-343
Arabs
256
Aral Sea
135
Arbenz, Jacobo
16
"Archangel-1"
262
"Archangel-2"
262
Area 51,
56, 72, 74-75, 77-78, 130-131, 133, 274, 283-284, 286, 291, 303, 305, 341
Arizona
80
Arkansas
78
Armenia
124, 144, 146
arms limitation negotiations
147
arms race
318
Army Air Corps
21, 56
Army Air Force
50, 328, 330
article 121
286, 288
article 126
296
article 133
296
article 341
66, 71, 131-132

article 345A
79
article 346
80
article 347
104
article 349
219
article 354
80
article 357
80
article 360
175, 217
article number
59
articles
20, 59, 99, 216
Ashford, Douglas E.
30
Asia
135, 190, 198, 211, 216, 221, 230, 233, 300-301, 310, 319
Asian Photographic Interpretation
229
aspheric lens
53
ASPIC
229
Aswan Dam
112
atmosphere
278, 25, 152, 182, 215
atmospheric testing
147
atomic bomb
19
Atomic Energy Commission
56, 133, 182, 241-242, 274
atomic energy facilities
233
ATR (Air Transport Rating)
187
Atsugi, Japan
134, 215, 218
autopilot
76, 341-342
Aviation Week
216
AXILLARY
342
Ayer, Frederick, Jr.
14-15, 36, 327

B

B model camera
53
B-17 bomber
222
B-26 bomber
214
B-36 bomber
22, 50
B-47 bomber
72
B-52 bomber
20, 111, 291, 312
B-57 bomber
5, 8-9, 13, 15, 24, 35, 72
B-58 bomber
262, 269-270, 273
B-58 "Hustler" bomber
262
B-58A bomber
269
B-58B "Super Hustler"
262
B-58B bomber
263, 269
B-70 bomber
279, 293
BACKFIN
167
"badlands"
71
Baikal, Lake
128
Baird, Walter
56
Baird Associates
56
Baker, James G.
18, 21-25, 29-31, 46, 50-56, 108, 253, 282, 327, 332-333
Balkhash, Lake
138, 139
ballistic missiles
2, 201, 206, 299, 318

25X1, E.O.13526

Baltic Sea
3, 303
Baltic states
104
Baltimore, Maryland
6, 9
Baotou, China
233, [redacted] 239
Barents Sea
140, 142, 172
Barnes, Tracy
101, [redacted]
barrel rolls
76
Batlin, Alex
66
Bay of Bengal
231
Bay of Pigs
191, 197-199, 205
Baykonur, Soviet Union
137
Beacon Hill
17-19, 21, 24, 31-32, 331-332
BEACON HILL Report
17, 19, 21, 24
BEACON HILL Study Group
18, 21
Beechcraft
56, 187
Beerli, Stanley W.
192, 196
Beijing, China (see Peiping, China)
228-229
Bell Aircraft Corporation
9
Bell Laboratories
18
Bell Telephone Laboratories
29
"bends"
64
Bennington, Herbert
309
Benton, William
299
Berezovskiy, Soviet Union
138
Berlin, Germany
16, 20, 127, 147, 162, 183, 196
Berlin crisis
147, 196
BGMARQUE
222
bhangmeter
240
BIRDWATCHER
235
Bison bomber
20-21, 27, 31, 98, 105, 111
Bissell, Richard M., Jr.
15-16, 30, 39-41, 43-44, 56, 59, 61, 66, 69-70, 72, 74, 78, 81-82, 89, 93, 95, 100-101, 104-106, 109-110, 121-122, 124, 127-128, 135, 139, 148, 154-157, 162, 168, 170, 172, 174, 178-179, 187, 191-192, 196, 214, 216, 224, 259-260, 262-263, 267, 277-278, 283, 285, 286-287, 289, 295, 327, 328-329, 330-331
Bissell's Narrow-Gauge Airline
72
"black books"
188
"black" contracts
285
BLACK KNIGHT
126, 134
Black Sea
85, 126, 133, 144, 153
BLACK SHIELD
301
Blip/Scan Ratio
267-268
BLUE BOOK
72-73
Board of Inquiry
184
Bodo AFB, Norway
142, 172, 176-177
Boeing RB-47
3
Boeing KC-135
71
bomb damage
49
bomber
2-3, 5, 20-23, 26, 31, 50, 66, 98, 105, 109, 111, 140, 167, 214, 262-263, 269, 279, 293, 316-317
Bomber Gap
31, 111, 316

borax
121
Boston, Massachusetts
13, 18, 22, 24, 27, 52, 56, 110, 260
Boston camera
13
Boston University
22, 24, 52
Bowles, Chester
225
Brahmaputra Valley
231
bridges
82
Briggs School
42
Briglia, Frank
339
British and French fleets
116
British Guiana
211
British Guiana-Venezuela border
211
British overflight of the Soviet Union
23-24, 164
British Parliament
94, 232
British radar network
94
Brugioni, Dino
135, 137
Budapest, Hungary
123
Budget, Bureau of the
267, 308-310
Budget, Director of the
40
Buffalo, New York
9, 63
Bulganin, Nikolai
94, 120
Bulgaria
101, 124-125, 140
Bundy, McGeorge
188, 201, 205, 209, 236, 293
BUORL
22
Burbank, California
45
Burke, William
168, 170, 174
Burma
216, 231, [redacted]
Byelorussia
104, 106

25X1, E.O.13526

C

C model camera
54
C-118 aircraft
144
C-124 cargo planes
66, 135, 181, 219
Cabell, Charles P.
30, 36, 43, 88, 101, 122-123, 127, 139, 183, 223, 295, 327, 329
Cairo, Egypt
117, 119
Calcutta, India
232
California
9, 13, 24, 26, 36, 45, 50, 54, 56, 133, 182, 198, 200, 208, 221, 247, 251-252, 254-256, 295, 311, 340
California Institute of Technology
26
Cambodia
216, 230, 243
camera lenses
93
camera technicians
78
camera-carrying balloons
14, 18, 23, 84, 145
cameras
1, 3-4, 12-13, 21-22, 31, 44, 49-51, 53-55, 60, 77, 87, 196, 244, 281-282, 310, 313, 321, 327, 329, 332, 333
Canberra
5, 9, 23-24, 67, 109, 117, 162
Cannon, Clarence
88
Capp, Al
45
cardiovascular system
62

25X1, E.O.13526

Carey, Howard
80, 140
Carter, John H. (Jack)
9-10
Carter, Marshall S.
186, 201, 205, 208-209, 299, 327
Caspian Sea
126
catheterized
64
cathode-ray tube
267
Caucasus Mountains
179
celestial "fixes"
76
Central Asia
135
Central Asian republics
176
Central Intelligence Agency (CIA)
29, 32, 59
Air Force competition
308
air operations
157, 191
General Counsel
44
Inspector General
190
proprietary
260, 129
proprietary firm
110, 129, 222, 260
U-2 program
257, 181
centrifugal force
177
CHALICE
155, 170, 182
Chamoun, Camille
152
Chance Vought Corporation
8
Ch'ang-sha, China
228
[redacted]
Charbatia, India
232-233, 237, 239
Charyk, Joseph V.
260, 327-328

25X1, E.O.13526

[redacted]
Chekiang Province
235
Chelyabinsk, Soviet Union
176
Cherbonneaux, James
127, 138, 214
Chiang Ching-kuo
224
Chiang Kai-shek
224
Chiang-yu, China
228
Chicago Daily Tribune
132
Chief of Station for Germany
101
[redacted]
Chief of the German General Staff
1
"chimney"
71
China
5, 85, 135, 148, 157, 174, 190, 201, 211, 215-217, 219, 221-222, 224-226, 228-233, 235-237, 239-240, 242-243, 245, 247, 253, 282, 301, 305, 307, 319, 336, 339-[redacted] 343.
See also Nationalist China and People's Republic of China (PRC)
China Lake, California
340
chine
277
Chinese Communist nuclear tests
240
Chinese mainland
319, 215, 222, 226
Chinese nuclear reactors
238
Chinese nuclear test
237
Christ, David L.
339
Christian Science Monitor
18
Civil Air Transport
214, 222
civilian supersonic transport (SST)
293

CL-282 Project
10-17, 24-27, 29-37, 66, 327-331, 333
CL-400 aircraft
263
clandestine operations
16, 321
Clandestine Services
192
Clark Airfield, Philippines
212
Clifford, Clark
311
Cline, Ray
209, [redacted]
"coffin corner"
76
Cold War
17
collection requirements
80-81, 189, 253, 300
Columbia Broadcasting System
18
combat aircraft
6, 26, 45
Combined Shipping Adjustment Board
15
Commerce, Department of
254
commercial airliners
72
communications intelligence (COMINT)
189, 311, 335-336
Communications Satellite Corporation
328
COMPASS TRIP
254
competitive bidding
44
compressor vanes
291
computer-controlled radar
294
computers
52
Congress, United States
44, 88, 98, 111, 159-160, 170, 293, 315, 317-318
constant-velocity film transport
281
Constellation airliner
10, 329
Contingency Reserve Fund
40, 44
contractors
42, 72, 259, 288
contrails
152
Convair
262-263, 267-271, 273, 284, 312
Convair F-106
284
convective clouds
94
Cornell Aeronautical Laboratory
18, 24-25, 328
Corning Glass Works
282
cosmic-ray effects
94
cover officer
89
covert activities
2, 6, 40, 43-44, [redacted] 80, 88, 134, 157, 189, 191-192, 212, 214, 222, 320-321, 328, 342
Covert Action Staff (CAS)
214
covert funding
43
Crabb, Lionel
94
Crete
116
Crick, Alan
154
Crick Committee
154
Crimean Peninsula
108
Cuba
188, 197-201, 205-211, 299-300, 302, 310, 327, 339
Cuban air and ground order of battle
198
Cuban Missile Crisis
55, 199, 209-210, 299, 327
Cubi Point Naval Air Station, Philippines
212, 216, 221
Culbertson, Allman T.
66
Cunningham, James A., Jr.
40, 66, 69, 72, 78-79, 125, 247, 288-289, 328-329
Current Intelligence, Office of
81

25X1, E.O.13526

25X1, E.O.13526

25X1, E.O.13526

Curtiss-Wright Corporation
328
Cyprus
[redacted] 257, 114, 116-117, 120
Czechoslovakia
101

D

D-21 aircraft
291
Dalai Lama
216
Dallas, Texas
293
Damage Assessment Team
183
David Clark Company
62
Davis, Saville
18
Dayton, Ohio
4, 8
DC-6 aircraft
341
deep penetration overflights
316, 123
Defector Reception Center (DRC)
2
defectors
2
Defense, Department of
33, 61, 122, 145, 154, 161, 234, 236-237, 239, 267, 292-293, 301-302, 304, 310, 317
Defense, Deputy Secretary of
81, 127, 144, 161, 188, 207, 300, 309
Defense, Secretary of
161, 298
Defense Mobilization, Office of
17, 26
Science Advisory Committee
37
Del Rio, Texas
256, 198
Delta-II camera
250
denied territory
174, 188, 190
Deputy Director of Central Intelligence (DDCI)
30, 36, 43, 327, 329, 88, 101, 122-123, 127, 139, 186-187, 201, 208-209, 223
Detachment A
94-95, 100-101, 106, 114, 116, 124-125, 139-140, 142-143, 181, 316
Detachment B
113, 116, 120-121, 124-126, 133, 135, 139, 142-143, 152-153, 155-157, 164, 167, 181-182, 187, 316
Detachment C
133-134, 139, 143-144, 152, 157, 182, 211-212, 214-220, 224, 316
Detachment G
133, 181-182, 198-199, 211, 221-222, 230-233, 243, 248, 250-251, 254-257
Detachment H
225-226, 228-230, 232-237, 239-241, 243-246, 250-251, 255
Development, Deputy Chief of Staff for
7, 11, 17, 35
Development Planning, Office of
12
Development Projects Division (DPD)
157, 191
Development Projects Staff
40, 55-56, 132, 157, 191, 212, 331
Dirty-Birds
128, 130, 135
Donovan, Allen F.
18, 21, 24-26, 29-31, 35, 260, 328
Donovan, James
183
Doolittle, James H.
61, 167, 328
Douglas Aircraft
339
DRAGON
2
DRAGON LADY
77
drivers
59
drone
291, 343
Dryden, Hugh
89, 182
DuBridge, Lee
26-27

Dulles, Allen
16-17, 30, 32-37, 39-40, 43-44, 56, 60, 73, 76, 80, 82, 86, 88, 93, 95, 97, 105-106, 109-111, 117, 120, 124, 127, 154, 161-162, 164-165, 167, 170, 180, 184, 189, 191, 214, 223-224, 260, 263, 295, 316-318, 327-328
Dulles, John Foster
20, 86, 109, 117, 122-123, 126
dummy corporations
44
Dunaway, Glendon
65

E

Earman, J. S.
206
earth's atmosphere
278
East Berlin riots of June 1953
16
East Germany
100-101, 142, 147, 183, 303
Eastern Europe
2, 84, 88-89, 100-101, 106, 110, 123-124, 126-127, 139-140, 142, 144, 304, 316
Eastman Kodak
4, 31-32
eavesdropping
2
EBONY
222
Eden, Anthony
93-95
Edens, Buster
135
Edgerton, Germeshausen & Grier (EG&G)
130, 274
Edwards AFB, California
133, 149, 181, 198, 200, 211, 230
Eielson AFB, Alaska
133-134
Eisenhower, Dwight D.
19, 26-27, 33-37, 39, 44, 56, 60, 73, 85-86, 88, 94-101, 104-106, 109-110, 112, 114, 116-117, 119-120, 122-124, 126-129, 135, 139, 143-147, 152-156, 159-164, 167-168, 170-172, 180-181, 187, 191-192, 195, 211-212, 214, 221, 224, 263, 267, 270, 277, 297, 315-317, 322, 328-330
electromagnetic energy
335
electromagnetic pulse indicator
240
electronic countermeasures (ECM)
207, 239, 304, 313, 317, 335-336
electronic intelligence (ELINT)
114, 126, 144, 146, 153, 162-163, 175, 182, 215, 242, 244, 303, 335
electronic intercept
126, 317
electronic sensors
292
electronic warfare
300
ELINT collection
142, 339
ELINT emanations
304
ELINT-collection gear
310
ELINT-collection unit (System VI)
170
Emergency Preparedness, Office of
254
engine nacelles
291
engineers
9-10, 12, 17, 25, 43, 45, 55, 61, 71, 255, 260, 269, 286, 339, 343
English Electric Company
5, 23
epoxy resin
277, 289
EQUINE
82
Ericson, Robert J.
80
espionage
1
Estonia
303, 339
Eurasian landmass
145
Europe
2, 4, 20, 23, 84-85, 88-89, 93, 100-101, 106, 110, 115, 123-124, 126-127, 139-140, 142, 144, 148, 178, 255, 304, 316, 330

Evang, Vilhelm
142
EVEN STEVEN
256
Executive Branch
109
exotic fuels
263

F

F-100 fighter
71
F-102 fighter
270
F-104 fighter
10-11, 45, 329
F-106 fighter
270, 284
F-108A Rapier
285
F-111 fighter
312
Fairbanks, Alaska
134
Fairchild K-19 aircraft
49
FAN SONG acquisition radar
234
Far East
3, 110, 133-134, 140, 144, 190, 197, 201, 216, 224, 233-234, 242-243, 301, 305, 316, 320, 332,
"Fast Move" operations
216
Federal Aviation Administrator
289
Federal procurement cycle
44
"ferret" equipment
3, 335
ferret flights
123
FFD-1
238
FFD-2
238-239
FFD-4,
282, 238
FICON
22
Field Activities, Deputy for
192
5412 Committee
188, 226
fighter
4, 10, 14, 22, 66, 74-75, 108, 124, 146, 177, 182, 201, 214, 312, 317, 328-329
fighter conversion
22
Fili airframe plant
105
fillets
277, 288-289
film
1, 31-32, 49, 51-53, 55, 60, 100, 108, 114-115, 119-121, 145, 153, 179, 209, 215, 218, 221-222, 226, 235-236, 239, 241, 247, 249-250, 254, 281-282, 306, 310-311, 316, 330
film processing
60, 121, 247
film-transport system
55
Firewel Company
63
First Weather Reconnaissance Squadron, Provisional (WRSP-1)
94
Fischer, C. W.
309
FISH
250, 263, 268-269, 312
FISH HAWK
249-250
Flickinger, Donald D.
62
FLIPTOP
198
"Flit"
62
Florida
198
flying boat
270
focal-length cameras
13
focal-length lenses
51

Follow-On Group (FOG)
77
Ford Foundation
15
Foreign Intelligence Advisory Board
192, 226, 292, 302, 330
forest fire detectors
238
Formosa
305 (See also Taiwan)
Fort Worth, Texas
50
40 Committee
188, 257
forward processing center
212
Foster, John
343
France
113, 116-117, 120, 201, 249
Franco-Prussian War
1
Frankfurt, Germany
2
Free Europe Committee
88
French atomic test area
250
French Indochina
222
French Mediterranean
116
French Mystere
116
French nuclear developments
249
French Polynesia
249
Frogman Incident
94
Fujisawa, Japan
218
Fukien Province
226, 229

G

G-forces
79
G2A aircraft
263
Galbraith, John Kenneth
231-232
Gardner, Trevor N.
11, 13-15, 26-27, 32, 36, 40, 66, 81, 99, 327, 329, 331
Gates, Thomas
164
Gaza Strip
114
Geary, Leo P.
61, 66, 71, 73-74, 78, 152, 186, 196, 288, 329
General Dynamics
312
General Electric J73/GE-3
11
General Tire and Rubber Company
329
GENETRIX
84-85, 87-88
Geneva, Switzerland
96, 147, 159
Geneva Conference on Surprise Attack
159
Geneva summit conference
96
geological survey
254-255
German Luftwaffe
22
German scientists
2
Gettysburg, Pennsylvania
120
Gibbs, Jack A.
61, 259-260
Giebelstadt Airbase, West Germany
95, 124, 140
Gilpatric, Roswell
207-208
Glennan, Keith
182
gliders
11
Glienecke Bridge
183
Globke, Hans
101
Goddard, George
330

Goering, Hermann
66
Goldmark, Peter C.
18
golf balls
50
Goodpaster, Andrew J.
89, 97, 100, 105-106, 109, 120, 126
128, 144, 146, 167, 172, 187, 329
Grace, Frank G.
80
GRANDSLAM
170, 172, 174-175
Granger Company
336
GRC-127
205
Great Britain
3, 7, 23, 93-95, 112-113, 115, 117,
120, 153-156, 181, 303, 319
Air Ministry
94, 155
counterintelligence operative
94
Greece
114
Greenland
172, 174
Greyhound bus
289
Groom Lake
56, 66, 133
Groton
15
ground resolution
281-282, 250
Guatemala
157
"guided democracy"
212
Gulf Coast
254
Gulf of Finland
142
Gulf of Tonkin Resolution
231
gust control
47
GUSTO
274

25X1 and 6, E.O.13526

25X1, E.O.13526

H

H camera
244, 247, 253, 256, 282
Hagerstown, Maryland
9
Haiphong Harbor
247
Halaby, Najeeb E.
289
Hall, William
117, 119, 143
Haller-Raymond-Brown (HRB)
335
"handing off"
87
Hanford, Washington
238
"hanger queen"
175
Harvard University
18, 21-22, 29, 50-51, 327, 129
Harvard College Observatory
21
Hawk-750 glider kit
342
haze filters
105
HBJARGON
175-176
HBJGO
176
heavy water
19
Helms, Richard M.
253, 256, 304-305, 309-310, 329
HEN HOUSE
168
HEN ROOST installations
168
Herter, Christian
127, 163
"hiding in the open"
55
high-acuity cameras
21, 50-51, 327
high-altitude air sampling (HASP)
219
high-altitude aircraft
4, 8, 16, 24, 59, 148, 304

high-altitude balloon program
22
high-altitude photography
1, 55, 84
high-altitude photoreconnaissance
7
high-altitude reconnaissance
1, 4-5, 8-9, 13-14, 17, 19, 22-25, 33-34, 39, 49, 52, 55, 254, 262, 292, 312, 330
high-altitude weather plane
178
high-impact plastics
313
high-resolution cameras
4
high-resolution infrared scanner
238
high-speed buffet
76
high-speed reconnaissance aircraft
262-263, 270
HIGH WIRE
164
Hilsman, Roger
209
Hitchcock, Alfred
29
Hokkaido, Japan
4
Holcomb, William
285
Hollywood, California
29
Hong Kong
217
Hoover, Herbert, Jr.
123-124
horizontal stabilizers
47
Horner, Richard
70
Hornig, Donald F.
309
HOT SHOP
162
Houston, Lawrence R.
44
HR-333 camera
253
HRB-Singer Corporation
336
HTAUTOMAT
83, 145, 334
HTNAMABLE
110-111
human intelligence sources
2
Hungarian uprising
123
Hungary
101, 123
Hurricane Baseline Survey
254
Hycon Manufacturing Company
13, 50, 252
hydraulic fluid
279
hydraulic pump
279
hydrogen bomb
19-20, 242
hydrogen weapons
164
hydrogen-powered aircraft
263
hydrologists
254

I

IBM CPC (card-programmed calculator)
52
IDEALIST
182, 210, 234, 303
IDEALIST U-2
303
Igarka, Soviet Union
3
ileitis
96
imagery
116, 179, 252-253, 255-256, 282, 298, 303, 310
impeller
291
"in-and-out" operations
205
in-flight refueling
198-200, 237, 302

Incirlik Airbase, Turkey
113, 181
India
148, 231-232, 237, 247, 319
Indian Parliament
232
Indian Springs, Nevada
274, 130, 132
Indonesia
211-212, 214, 319
Indonesian Air Force
212
inertial guidance systems
313
inflatable aircraft
262
infrared camera
282
infrared jammer
336
infrared scanner
238-239
inland waterways
82
instant-film camera
330
intelligence, collection of
16, 31-32, 80-81, 142, 146, 153, 162, 170, 182, 189, 211, 245, 253, 256, 282, 300, 303, 310, 315, 319, 322, 333, 336, 339
Intelligence and Research, Bureau of
209
intelligence community
2, 20, 33, 81, 138, 153, 160, 164-165, 170-171, 190, 196, 232-233, 254, 298, 303-304, 322, 339
Intelligence Coordination, Office of
81
intelligence requirements
25-26, 80-82, 114, 122, 154, 188-190, 226, 230, 233, 250, 253, 300, 332
Intelligence Requirements Officer
81
Intelligence Star
186
Intelligence Systems Panel (ISP)
21-22, 24-26, 30-32, 35, 327
intercept
87, 108, 124, 126-127, 142, 148, 162, 168, 170, 174, 176-177, 259, 285, 317, 335
interceptor aircraft
93, 109, 124, 148, 176
intercontinental ballistic missile
137, 159
Interior, Department of the
254
international date line
134
International Geophysical Year
85
international law
110
interrogation
2-3, 183, 241
intruder
109
"invisible" dirigible
18
ion cloud
300
ionosphere
175
Iran
124, 133, 189
Iraq
133, 153
Iron Curtain
1, 17, 93, 95, 111, 126, 146
ISINGLASS
312
Isle of Pines
205
Israel
113-114, 116-117, 120, 343
Israeli paratroop
117
Israeli-Egyptian conflict
257, 117
Israeli-Syrian disengagement
257
Itek Corporation
331, 250
IX spaceflights
254

J

J-3 (Operations)
189
J75 power plant
152
jamming devices
300
Jammu and Kashmir
231
Japanese airspace
4
Japanese Government
134, 182, 221
jet stream
94, 145
jet-propelled glider
11, 25
Johnson, Clarence L. (Kelly)
10-12, 13-14, 17, 24-26, 29-31, 34-36, 43-48, 50, 54, 56-57, 59, 64, 66-70, 76, 108, 130, 132, 149, 179, 185, 196, 247-248, 251, 262-263, 268, 271, 274, 277-279, 281, 284, 289, 293-297, 320, 329
Johnson, Lyndon B.
195, 236-237, 243, 293-295, 302, 304-305, 310-311, 332
Joint Chiefs of Staff (JCS)
122, 124, 126, 161, 187-190, 223, 234, 302
Joint Intelligence Committee
154
Joint Priorities Committee
154
Joint Reconnaissance Center (JRC)
189
Jones, Edwin K.
138
Jordan
117, 153
JP-7 fuel
62

K

K-21 framing cameras
49
K-38 camera
50-51, 53
Kadena Airbase, Okinawa, Japan
301
Kaliningrad missile plant
105
Kamchatka Peninsula
128, 134, 197
Kandalaksha, Soviet Union
176
Kapustin Yar, Soviet Union
16, 23-24, 109, 139, 143, 164
Kara Sea
142
Kazakh language
135
Kazakhstan
135
Kazan', Soviet Union
167
KC-135 tankers
71, 199
KEDLOCK
285
"Kelly's Angel"
66
KEMPSTER
300
Kennedy, John F.
186, 191-192, 195-196, 198, 207-208, 210, 225, 229, 232, 235, 292-293, 297-298, 330
Kennedy, Joseph W.
29
Kennedy, Robert
215
Kennedy administration
191, 195, 225
kerosene fuel
62
Khan, Ayub
135

Khimki rocket-engine plant
105
Khrushchev, Nikita
94, 96, 100-101, 147, 160, 163-164, 168, 179-180, 299
Kiangsi Province
235
Kiangsu Province
229
KICK OFF
198
Kickapoo Joy Juice
45
Kiefer, Eugene P.
7, 9, 11, 260, 330
Killian, James R.
27, 29-30, 33-34, 89, 100, 191-192, 263, 292, 330
KINGFISH
270, 273

[redacted]

25X1 and 6, E.O.13526

Kirkpatrick, Lyman B.
190
Kirov, Soviet Union
172, 176
Kirtland AFB
79
Kissinger, Henry
256
Kistiakowsky, George
171
Klein, Burton
11
Klyuchi, Soviet Union
134, 143

[redacted]

KNIFE EDGE
167
Knowles, Harold E.
344
Kohler, Foy D.
112
Koko Nor
233
Kola Peninsula
182

Withheld under statutory authority of the Central Intelligence Agency Act of 1949 (50 U.S.C., section 403g)

Kollman Instrument Company
59
Koon, Ralph E.
79
Korea, demilitarized zone (DMZ)
305 (See also North Korea)
Kotlas, Soviet Union
172
Krag, Jens Otto
299
Kratt, Jacob
65, 79, 116
Kremlin
123
Kucera, George
40
Kunming, China
226
Kunsan, South Korea
239
Kuybyshev, Soviet Union
164
KWCORK
113
KWEXTRA-00
59
KWGLITTER-00
59
Kyshtym, Soviet Union
176

L

L'il Abner
47
L-pills
66, 125
Lahore, Pakistan
135
Lakenheath AFB, England
93
Lan-chou
226, 233, 237-239

Land, Edwin H. (Din)
18, 21, 25, 29-34, 36, 80-81, 89, 96, 100, 110, 191-192, 260, 262-263, 269-271, 292, 307, 330, 332-333
Land Panel
37
Langley, Virginia
43, 192, 306,
Langley AFB, Virginia
23
Laos
216, 221, 231, 233
Las Vegas, Nevada
284
Latham, Allen, Jr.
29
Latvia
303
Laughlin AFB, Texas
198, 200
Lebanon
114, 117, 152-153, 343
Ledford, Jack C.
207, 288, 297
Leghorn, Richard S.
4-8, 18, 23, 110, 330
LeMay, Curtis E.
11-12, 14, 60
"Lemon-Drop Kid"
124
Lend-Lease
7
Leningrad, Soviet Union
104-106, 109, 142, 298, 303
lens
1, 18, 21-22, 30, 49-55, 93, 108, 250, 247, 253, 281-282, 321, 327
lens curvatures
52
lens designs
52, 55
LeVier, Tony
56, 68-71, 75
LF-1A fuel
62
Libya
121
lightweight film
32
LINCOLN
18
Lindsey Air Force Base, Wiesbaden, West Germany
80
liquid-fueled missiles
161, 317
lithium deuteride
19
Lithuania
303
Little, Arthur D., Inc.
29
Little Quemoy
215
Lloyd, Selwyn
93
Lockheed Aircraft Corporation
9, 24, 329
Lockheed Day Fighter
14
Lockheed mechanics
66, 130, 225
London School of Economics
15
LONG GREEN
198
long-range bombers
20, 81, 98-99, 316
LONG SHAFT
245
Lop Nor test site
237
Los Angeles, California
186, 254
Lovelace, W. Randolph, III
62, 74
Lovelace Foundation for Medical Education and Research
74
low-radar-cross-section vehicle
344
low-level photography
209
low-vapor-pressure fuel
278
Luke AFB, Arizona
74

Luna 2 rocket
163
Lundahl, Arthur C.
82, 114-115, 117, 119, 121, 135, 142, 154, 191, 209, 241, 331

M

M Building
83
M-12 aircraft
291
M-195 aircraft
9
MacArthur, John
167
Macdonald, Duncan E.
22, 154-155
Macmillan, Harold
155-156, 164, 167, 187
Malaya Sazanka, Soviet Union
144
Malinovsky, Rodion
160
Malta
114, 116
Manchuria
229, 233, 237
Mandrake
148
Manhattan Project
329
Manila, Philippines
311
manned space program
62
March AFB, California
75
Mark-30
335
Marquardt ramjets
268
Marr, William H.
72
Mars
79
[redacted]
Marshall Plan
15, 327
Martin, Glenn L., Aircraft Corporation
5, 8-9, 24, 270
Massachusetts Institute of Technology (MIT)
4, 15, 18, 27, 30, 110, 129-130, 327, 330, 332-333
Matomic Building
42, 104
MATS
72
Matsu Islands
229
May Day
20, 176, 191, 195
MC-2 and MC-3 partial-pressure suits
64
McConaughy, Walter P.
257
McCone, John A.
184-186, 192, 195, 197, 201, 206-207, 209, 232-234, 236-238, 240, 292-294, 296, 298, 300, 327, 331
McCoy, Frederick
95
McCullough Corporation
339
McDonnell Aircraft
312
McElroy, Neil
146, 161-162
McFadden, William
54
McMahon, John N.
186
McNamara, Robert
229, 234, 236, 238, 253, 285, 292-293, 298, 300-301
mechanical fuel totalizer/counters
68
Mediterranean Sea
114, 116, 121, 140, 142, 153
medium-range ballistic missiles
201, 206
Melpar, Incorporated
342
Menderes, Adnan
112, 181
mercenary pilots
214
metal fatigue
242

Withheld under statutory authority of the Central Intelligence Agency Act of 1949 (50 U.S.C., section 403g)

meteorologists
42, 87
MI-6
93, 154
Michigan, University of, School of Aeronautics
329
microwave reconnaissance
19
Middle East overflights
114, 116
MiG aircraft
336
MiG-15
108, 148
MiG-17
5, 108, 148
MiG-19
148
MiG-21
148, 237, 244
Military Adviser to the President
188
Military Air Transport Service
72
MILLTOWN
216
Miller, Stewart E.
18, 21, 39-40, 55-56, 331

[redacted] 25X1, E.O.13526

"million-dollar" photography
111
miniaturization
339
mirrors
54-55
Mirza
135
"missile gap"
157, 159, 317
missiles
2, 22, 81, 98-99, 137, 139, 149, 160-161, 165, 167, 171, 201, 205-206, 209, 226, 239, 243, 257, 285, 299, 301, 303-304, 306-307, 315, 317-318, 336, 339
Mississippi River
78, 310
MKNAOMI
66
Monthly Activities Book
189
moon
163
Morotai
214
Moscow, Soviet Union
3, 20, 85, 94, 101, 105-106, 109, 124, 143, 145, 160, 167, 177, 179, 318
MUDLARK
178

[redacted]

multiengine aircraft
12, 24-25
multiple sensors
22
Murmansk, Soviet Union
140, 142, 172, 174, 176
Mururoa Atoll
250
Myasishchev-4 (see Bison bomber)

Withheld under statutory authority of the Central Intelligence Agency Act of 1949 (50 U.S.C., section 403g)

N

Nagy, Imre
123
Nanking, China
228
Narcotics, Bureau of
254-255
Nasser, Gamal Abdel
112, 153
National Advisory Committee on Aeronautics
89, 94
National Aeronautics and Space Agency (NASA)
178-179, 254
National Air & Space Museum
64
National Intelligence Estimates
19, 235
national intelligence objectives
81
National Security Agency
81, 114, 134, 185, 327
National Security Bureau
224
National Security Council
27, 294, 161, 167, 187-188, 191, 197, 212

National Security Council Intelligence Directive (NSCID)
191
Nationalist China
319, 215, 224, 226, 229, 236
Nationalist Chinese
201, 205, 222-226, 229-230, 234-237, 239-240, 243, 245-246, 253, 257, 319
Air Force
223, 225
insignia
225
pilots
319, 223, 230, 235, 243, 246, 253
U-2
205, 223-224, 229, 246
Navajo missile
22
Naval Air Station at Atsugi, Japan
134
naval quarantine
210
Nehru, Jawaharlal
231-233
Nelson, Norman
68
Nevada
56-57, 341, 113, 130-131, 133, 149, 182, 241, 274, 283, 286, 290, 295, 297, 302-303, 311
Nevada Proving Ground
56, 241
New Delhi, India
232
Ngo Dinh Diem
230-
NICE GIRL
310
nickel-cadmium batteries
240
Nile River
119
NIMBUS
200
1973 Middle East war
256, 320
9th Tactical Air Force
330
nitrogen
64
Nixon, Richard M.
245-246, 256-257, 333
Nobel Prize
332
North American Air Defense Command (NORAD)
289
North Atlantic Treaty Organization (NATO)
20, 329, 87, 148, 172
radars
87
North Island Naval Base, California
248
North Korea
4, 233, 237, 242, 307
North Vietnam
190, 221-222, 230-231, 233, 243, 246-247, 301, 304-307, 311, 336, 342
North-East Frontier Agency (NEFA)
231
Northrop Aviation
259
Norton, Garrison
14-15, 31, 70, 260, 331
Norwalk, Connecticut
50
Norway
140, 142, 172, 176-177, 303, 319
Norwegian Defense Staff
142
Norwegian Intelligence Service
142
Notre Dame University
7
Novaya Zemlya
172
Novokuznetsk, Soviet Union
138
nuclear energy
39, 81, 225
Nuclear Energy Division
39

Withheld under statutory authority of the Central Intelligence Agency Act of 1949 (50 U.S.C., section 403g)

nuclear fireball
240
nuclear resonance
332
nuclear strike
292
nuclear testing moratorium
317
nuclear tests
133, 147, 240
nuclear weapons
2, 19-20, 99, 147, 226, 238, 240, 317

O

OARFISH
44
O'Donnell, Emmett (Rosy)
74
Oak Ridge, Tennessee
238
oblique photography
135, 189, 244
Offshore Islands Crisis
215-216
[redacted]
OILSTONE
60-61
Okinawa
301-302, 305[redacted]308, 311
Old Executive Office Building
27
OLIVE HARVEST
257
Omaha, Nebraska
60
112A camera
250
112B camera
250, 252
"Open-Skies"
96
Operation Overflight
186
optical distortion
282
Optical Research Laboratory
21-22
optical systems
52
order-of-battle
82, 134, 322
"Oscar-Sierra" unit
234
[redacted]
over-the-horizon radar
241
Overhage, Carl F. P.
18, 21, 331
overhead photography
82, 87, 172, 237-238
overhead reconnaissance missions
188
Overstreet, Carl K.
100
OXCART
61, 73, 157, 191, 236, 238, 252-253, 259, 274, 277-313, 320-322, 327-333, 336, 339
oxygenation
62

25X1, E.O.13526

P

P-38 fighter
10, 329
P-40 fighter
328
P-51 Mustang
214
P-80 fighter
10, 45
P-80 Shooting Star
45
P2V7
222
P3A
222
Pacific Ocean
247
Padang, West Sumatra
212

25X1, E.O.13526

25X1, E.O.13526

Pakistan
128, 135, [redacted] 148, 163, 167-168, 170, 172, 174, 178, 189, 247, 316, 319
Paksane, Laos
231
Palmdale, California
311
panoramic-type framing camera
53
Paradise Ranch
57, 66
PARAMOUNT Committee
114, 116, 120
Parangosky, John
274, 288, 296, 309, 331
"parasite" aircraft
263
Paris Summit
172, 181-182
Pasadena, California
13, 50, 252
Pathet Lao
221
PB-4Y aircraft
222
Pearl Harbor
322
Pearson, Drew
299
Peenemunde rocket project
2
Peiping, China
228-229
Pellegrini, Joseph J.
6
Pennsylvania State Police
222
Pentagon
5, 7, 9, 11, 13-14, 39, 78-79, 238
People's Liberation Army (PLA)
215
People's Republic of China (PRC)
85, 135, 201, 215, 222, 225-226, 228-247, 253, 282, 301, 319, 339, 341
(See also China)
radars
215
troop movements
215
nuclear plants
238
nuclear weapons establishment
226
Shuangchengzi missile-test range
235
periscope
56, 76
Perkin, Richard S.
18, 31, 50-52, 281-282, 311, 332-333
Perkin-Elmer (P-E) Company
18, 31, 50, 281
Perkin-Elmer tracking camera
52
Perkins, Courtland D.
260
Personnel (USAF), Deputy Chief of Staff for
74
Peshawar, Pakistan
135, 163-164, 167-168, 172, 174-176

[redacted] 25X1, E.O.13526

Photo-Intelligence Division (PID)
82, 100, 135
photographic chemicals
121
Photographic Evaluation of Information on Cuba
200
photographic resolution
281
photointerpreters
83, 108, 111, 114-116, 120-121, 135, 138, 153, 172, 191, 198, 209, 211, 216, 221, 236, 303, 307
photomapping
1, 232
photoreconnaissance
1, 4-5, 7, 22
photosatellite
270, 313, 327, 330, 157
Phouma, Souvanna
221
PID's Special Projects Branch
115
PINE RIDGE
343-344
"pioneer" photography
87
pitot-static tube
295
Pittsburgh-Corning Glass Company
55
Plans, Deputy Director for
157, 191-192, 327, 333

Plans, Directorate of
73, 81, 157, 197, 211, 287
Plesetsk, Soviet Union
176
Poland
100-101, 104, 110, 123, 145, 303
polarizing filters
330
Polaroid Corporation
18, 24, 29
POLECAT
221
Polish Ambassador to the US
109
Polyarny Ural
172
polygraph examination
185
Pope, Allen L.
214-215
poppy fields
254
Port Said, Egypt
120
Portsmouth Harbor
94
potassium cyanide
66, 125
powered glider
14, 25, 339
Powers, Francis Gary
88, 117, 124, 174-178, 181-187, 190-191, 195, 197, 201, 219, 221, 224, 252, 316, 322, 332, 336
Poznan, Poland
104
PR3 aircraft
5
Pratt, Perry W.
31
Pratt & Whitney
12, 31, 35, 40, 48, 67, 71, 152, 217, 251, 268, 270, 284-285, 290
Pratt & Whitney J58 power plant
270
Pratt & Whitney J75-P13 jet engine
152
Pratt & Whitney J75/19W engines
284
Pratt and Whitney J57
12
President's Board of Consultants on Foreign Intelligence Activities
37, 127, 167, 170, 190, 192, 328, 330
President's Foreign Intelligence Advisory Board
37
President's Intelligence Checklist
201
President's Science Advisory Committee
37
pressure suits
64, 247
pressurized "spacesuit"
63
Prettyman, E. Barrett
184-185
Prettyman Board
185
Princeton University
29
processing and interpretation centers
115
Project Director
61, 82, 329
Project Officer, AFCIG-5
61
Project Staff
40-43, 56, 72-73, 82, 89, 170, 299
Project Three
27, 29-34, 327
propaganda
88, 160, 162, 183, 186, 317
Pryor, Frederick
183
psychological warfare
186, 222
Pueblo
242, 307, 320
Puerto Rico
207, 211
Pulse-Position Indicator display
267
Purcell, Edward M.
18, 29, 110-111, 129, 260, 263, 271, 274, 332
Putt, Donald L.
11, 35-36, 40
PVO [Soviet Air Defense]
174

Pyroceram
268, 271

Q

Quarles, Donald
81, 127, 144-145, 161-162, 263, 267
Quarters Eye
42
quartz glass window
282
Que Building
83, 111
Quemoy
215, 229
QUICK KICK
216
quick-reaction capability
301, 313

R

Raborn, William F., Jr.
240, 301, 332
radar
3, 7, 15, 19, 23, , 82, 87, 94, 97, 101, 104-106, 110-111, 123, 128-130, 132-134, 139, 142, 144, 148, 168, 175-176, 178, 197, 211, 215, 230, 234-237, 239, 241-243, 247, 257, 259-260, 263, 267-271, 273-274, 277, 289, 294, 300, 303-307, 310-311, 317, 331-332, 335-336, 339, 341-342

radar camouflage
128
radar cross section
15, 260, 263, 267-270, 271, 273-274, 277, 300, 332, 339, 341-343, 110
radar deception
130
radar defense systems
343
radar image
259, 332
radar pulses
87, 110, 130, 175, 268, 335
radar signature
341-342
radar technology
7
radar-absorbing
111, 260
radar-deflecting
260
radar-homing unit
342
Radford, Arthur W.
122-124, 126
Radhakrishnan, Savepalli
232
Radiation Laboratories
330
RADINT
189
radio
2, 7, 19, 75, 87, 175-176, 186, 268, 289, 332
RAINBOW
111, 129-130, 259
Ramenskoye, Soviet Union
20, 105
Ramey AFB, Puerto Rico
211
ramjet-powered
262-263, 268, 291
Ramo-Wooldridge firm
335
RAND Corporation
20, 26, 85
Randolph AFB, Texas
73
Randsburg Wash
340
Ray, Walter
302
"ray-tracing" programs
52, 55
RB-29 aircraft
3-4
RB-47 aircraft
3, 22, 182, 186
RB-52 aircraft
22
RB-57 aircraft
26, 35, 40, 134, 223
RB-57D aircraft
6, 126, 162

RB-58 aircraft
22
Reber, James
81, 114-115, 154, 332
reconnaissance
1-9, 12-19, 21-25, 27, 30-37, 39, 41, 43, 46, 49, 52, 55, 59, 61-62, 66, 80, 84, 94, 96, 98, 109, 113, 115-116, 119, 122, 126-127, 133, 153-154, 161-162, 188-193, 196-197, 201, 205-206, 208, 210-212, 216, 222-223, 226, 228, 230-233, 240, 249, 253-254, 260, 262-263, 267-268, 270, 287, 291-292, 295, 297-299, 301, 303-304, 308-309, 311-313, 319-322, 327-331, 333, 339, 341-344
reconnaissance aircraft competition
9
reconnaissance drones
301
reconnaissance satellite project
161
"reconnaissance strike"
295
Reconnaissance Systems Branch
4
records, altitude
302
records, speed
302
RED DOT
254
reflecting cube
281
reflective system
55
refueling
114, 198-200, 231, 237, 241, 302, 306, 308
refugees
2
Republic RF-84F Thunderflash
22
Research, Directorate of
192, 287
Research and Development, Director of
66
Research and Reports (ORR), Office of
21, 81, 237
Returnee Exploitation Group (REG)
2
Revolutionary Council
212
Rhodes
116-117
Ridenour, Louis N.
18
Ridenour Associates, Inc.
18
Ritland, Osmond
56, 61, 70, 78
ROBIN
23
ROBIN HOOD
211
Robinson, Robert
164
rocket planes
60
rocket-powered aircraft
312
Rodgers, Franklin A.
129, 267-268, 274, 332
Rodgers' Effect
267
Rolls-Royce Avon-109 engines
5, 23
Romania
101
Root, L. Eugene
9, 24
Rose, Wilburn S.
40, 79-80
Royal Air Force (RAF)
5, 23, 93-94, 109, 143, 153-156, 164, 181, 187, [redacted] 303, 319
fighters
94
intelligence officials
23
officers
153, 155
runway
57, 68-69, 79-80, 108, 140, 176, 283-284, 291, 302, 305
weather service
156
RS-71
295
Ruseckas, Joseph
63
Rusk, Dean
183, 201, 209, 236, 293, 304, 307

25X1 and 6, E.O.13526

Russell, Richard B.
88
Ryan Aircraft
343

S

S-band radar
87
sailplane
10-11, 24-26, 47
Sakhalin Island
201
Saltonstall, Leverett
88
SAM site
177, 205, 306
San Cristobal, Cuba
206-207, 209
San Diego, California
247, 343
SANDY HOOK
343
Santa Barbara Channel
254
Santa Clara Airfield, Cuba
201
Saratov Engels Airfield, Soviet Union
164
Sary-shagan, Soviet Union
139, 168, 232, 336
satellite imagery
298, 255
satellite navigation systems
341
satellite photography
87, 195-197, 226, 233, 237, 241
satellites
22, 100, 104, 159, 162, 211, 299, 303, 312, 320, 322, 328
Saville, Gordon P.
17-18
Schalk, Louis
288-289
Schlesinger, James R.
257, 333
Schoech, William A.
285
Schriever, Bernard A.
7, 11
Schultze, Charles L.
309
Science and Technology, Directorate of
193, 287, 328
Science and Technology, Special Assistant to the President for
37
Scientific Advisory Board
17, 23
Scientific Advisory Committee
17, 89, 332
Scientific Engineering Institute (SEI)
110-111, 129, 260, 267, 274, 331-332
Scientific Intelligence (OSI), Office of
14, 21, 39, 81, 97, 331, 333
SCOPE COTTON
310
SCOPE CRANE
242
SCOPE LOGIC
303
SCOPE SAINT
255
SCOPE SAINT-I
255
SCOPE SAINT-II
255
SCOPE SHIELD
246-247
Scott, Roderick M.
51, 53, 281, 333
Scoville, Herbert
207, 287, 289, 333
SCR-270 radar
7
Scranton-Wilkes-Barre Airport, Pennsylvania
222
Sea of Japan
3
Seaberg, John
8-9, 12, 35-36, 333
SEAFOAM
211
sealant
286, 288
secrecy agreements
295

security
2, 27, 40-41, 43, 55, 59-60, 72, 81, 94, 112, 125, 127, 140, 161, 167, 185, 187-188, 191, 197, 201, 209, 212, 218, 221-222, 224, 236, 256, 273, 283-284, 294, 318, 327
Security, Office of
59, 112, 222
Semipalatinsk, Soviet Union
128, 138-139, 147, 168
SENSINT--Sensitive Intelligence
3
sensors
11, 22, 241, 292, 310
Seventh Air Division
89
Severodvinsk, Soviet Union
172, 176
Shanghai, China
222, 228, 242
sheep-dipping
74
Shell Oil Company
62, 328
shellfish toxin
125
Shelton, William
125, 176
shock wave
290-291
short-range reconnaissance vehicle
342
show trial
183
Shuangchengzi, China
226, 228, 235
shutter assemblies
53
Siberia
3
Sichuan, China
230
Sicily
116
Sieker, Robert
131
Sierra snowfield
254
SIGINT Committee
332
Signals Intelligence (SIGINT)
222
Sinai peninsula
117
"single-pass" mission
306
Sino-Indian border
231-233, 319
Sino-Soviet border
336
"Skonk Works"
45
Skunk Works
45, 71, 132, 251, 262, 273-274
SKYLARK
299
Smithsonian Institution
55
"snap-up"
148
Snark missiles
22
Snider, Sammy V. C.
138
snowmelt
254
SOFT TOUCH
135, 139, 143
software algorithms
52
solar cells
240
solid-fueled missiles
161, 317
sound barrier
289
SOUTH GATE
216
South Vietnam
230-231, 243, 301, 305
Southeast Asia
190, 198, 216, 221, 301, 310
Sovetskaya Gavan', Soviet Union
144
Soviet Bloc
1-2, 4, 6, 16-20, 87, 93, 112, 123, 147, 152, 159, 195, 197, 211, 330
Soviet Communist Party Congress
160
Soviet Far East
134, 144

Soviet Union
1-3, 4-7, 14-27, 31, 39, 65, 81-89, 93-101, 104-106, 108-112, 120, 122-124, 126-130, 133-135, 137-140, 142-149, 152-157, 159-165, 167-168, 170-172, 174-183, 186-187, 189, 195-197, 200-201, 205-206, 209-211, 215, 219, 221, 225, 232, 237, 247, 254, 259, 294, 297-299, 301, 303-304, 312, 315-320, 322, 329-330, 332, 335-336, 339, 341, 344
activity in Cuba
200, 201
air defense system
2, 3, 108, 168, 170, 317
air power of the
20
air show
98
atomic energy program
317
bomber
21, 111, 167, 317
borders
123, 127, 143
diplomatic protests
315
electronic defenses
317
fighters
3, 5, 23, 26, 127, 144
G-class ballistic-missile submarine
237
guided missile and nuclear programs
135
guided missile research
99
guided missiles
81, 99
guided-missile test range
16
intentions
4, 318
intercept attempts
142
interception of U-2
19, 24-25, 47, 87, 106, 108, 110, 122, 127, 134, 144, 170, 244, 270, 315, 339
medium-range ballistic missiles
201, 206
missile program
128, 159-161, 163-164
missile test range
16, 23, 143, 164, 167, 176
overflight operation
110
propaganda campaign
317
Protest Note
109
radar coverage
106
radars
3, 7, 15, 97, 130, 134, 168, 176, 259, 339
rail network
167
strategic capabilities
2, 19
Supreme Soviet
179
surface-to-air missile (SAM)
177, 168, 304
surprise attack
19, 26-27, 89, 147, 159, 329-330
Soviet-Finnish border
303, 142
Soviet-Iranian border
124, 162-163, 168
Special Activities, Director of
193
Special Activities (OSA), Office of
43, 192-193, 207, 234-235, 242, 247, 252, 257, 288, 296-297, 299, 303, 312, 328, 341-342, 344
Special Assistant for National Security Affairs
188
Special Assistant to the DCI
40
"special" missions
116
Special Operations Division
192
Spica, Incorporated
51
spotting camera
282
Sputnik
159
SQUARE DEAL
168, 170

SR-71
295, 302, 308-311, 321
SS-4 (SHYSTER) MRBM
206
SS-6 ICBM
160, 165, 171-172

[redacted] 25X1 and 6, E.O.13526

stainless steel
277, 280
Stassen, Harold
330
State Department
81, 95, 101, 146, 154, 163, 179, 186, 196-197, 212, 224-225, 229, 243, 249
"stealth" research
313
stereo camera
281, 252
stereoscopic lenses
250
Steuart Building
83, 111, 114
STEVER
260, 333
Stockman, Hervey
65, 104-105
STPOLLY
222, 224, 240
Strategic Air Command (SAC)
11, 14, 17, 20, 26, 42, 60, 74-76, 78, 89, 93, 95, 198, 201, 207-210, 299, 230
meteorologists
42
strategic bombing
49
"strategic hamlet"
230
strategic intelligence
82, 115, 152, 302, 304, 313, 315, 319, 322
strategic reconnaissance
9, 122, 230, 295, 297, 299, 313, 320, 322, 329
Strauss, Lewis
56
Strong, Philip G.
14-17, 21-22, 24-25, 29-30, 333
subsonic
8, 14, 260, 263
Suez Canal
112-114, 117, 120, 256
Suez Canal, Israeli-British-French attack on
116
Suez Canal Company
112-113
Suez Crisis
152
Suhrawardy, Huseyn Shaheed
135
suicide pill
65
Sukarno, Achmed
212, 215
Sulawesi (Celebes)
212
Sullivan, Dennis
307
Sumatra
212
SUN SPOT
171
Super Snooper
132
supersonic designs
260
supersonic interceptor
285
supersonic technology
292-293
Support, Directorate of
40, 83
surface-to-air missile (SA-2) sites
201
surface-to-air missiles
257, 299, 301, 303, 339
surface-to-surface missiles
205-206, 209, 304, 306
Sverdlovsk, Soviet Union
172, 176-179, 185
Symington, Stuart
99, 161
synthetic lubricants
279
Syria
114, 120, 133, 153
Syrian radars
343
System-I
335

System-II
335
System-III
335-336
System-IV
335
System-V
126, 215, 335
System-VI
335, 175
System-VII
335-336
System-VIII
335
System-IX
335
System-IXB device
175
System-X
256, 336
System-XII units
234
System-XIII
234, 236-237, 239
System-XVI
336
System-XVII
336
System-XX
336
System-XXI
336
System-XXII
336
System-XXIV
256

T

[redacted] **25X1, E.O.13526**

Taber, John
88
TACKLE
224, 226, 228-230, 234-239, 241-246, 2 53, 257, 282
tactical intelligence
304
tactical reconnaissance
9, 25, 115, 153
TAGBOARD
291
Tahiti
250
tailless subsonic aircraft
263
Taiwan
215-216, 222, 224-225, 229, 233, 236-237, 239, 241-243, 246, 251
Taiwan Strait
229
Ta Khli, AFB, Thailand
230
Ta Khli, Thailand
217, 231, 237, 241
Talbott, Harold E.
13-14, 36
TALL KING
294
Tallinn, Estonia
303
T'ao-yuan Airfield, Taiwan
225, 230, 237, 239, 243
Tashkent, Soviet Union
176
TASS
137
Tatar Strait
144
Taylor, Maxwell
188
TCP
27
Technical Services Division (TSD)
66, 125
Technicolor
331
Technological Capabilities Panel
26-27, 29-30, 35, 62, 80, 327-329, 332
Technology, Deputy for
192
Teheran, Iran
144
Teledyne J-100-CA-100 Turbojet engine
343
Teledyne Ryan Aeronautical Company
343

television reconnaissance package
343
Tennessee
78, 238
tent-ology
116
test ban
147
Tet offensive
243
Texas Instruments Corporation
238, 282
Thailand
217, 219, 230-232, 237, [redacted]
305-306, 319
Third World
211
three-camera trimetrogon unit
52
three-dimensional movies
29
30001 camera
333
303 Committee
243, 257, 301-302, 304, 307
Thule, Greenland
172
Tibet
157, 216, 219, 232
Time
99
TIME STEP
172, 174
Titan ICBM
160
titanium alloy
271
Titanium Metals Corporation
280
TOKEN
87
Tomsk, Soviet Union
138
TOPPER
219
TOUCHDOWN
163
Toulon, France
116
tracker camera
241
tracking
3, 7, 52-53, 87, 97, 106, 128-129, 144, 170, 176, 178, 234, 259, 263, 267-268, 303, 307, 343
tracking radars
7, 267
Trans-Siberian Railroad
144
transport aircraft
25, 45, 66
traveling-wave tube (TWT)
236
trimetrogon K-17 mapping-camera system
49
Tripartite Declaration, 1950
117
tropopause
152
Truman administration
17
TRW Corporation
335
Tuamotu Archipelago
249
"Tuesday lunch"
305
Tukey, John W.
29
turbojet engine
8, 11, 262, 268, 343
Turkey
3, 112-114, 116, 121, 135, 146, 152-153, 178, 181, 189, 247, 316, 319
Tuslog Detachment 10-10
113
Twining, Nathan
60-61, 86, 88, 96, 99, 101, 127, 134, 139, 161, 165
Tyler Building
43
Type-I camera
281, 304, 310
Type-II camera
281
Type-IV camera
282
Typhoon Winnie
216
Tysons Corner, Virginia
43

25X1, E.O.13526

Tyuratam, Soviet Union
135, 137-138, 160, 163-165, 167-168, 171-172, 176

U

U-1
66
U-2
accomplishments
316
assembly
45, 47, 66, 75, 164, 280, 284, 291
bailout experiments
64
bases
93
bicycle-type landing gear
47
blue-black color
149
brakes
69
canopy
80, 149, 177
cockpit
11, 63-64, 79, 125, 283, 292, 296
cost
34, 124, 149, 250-252, 263, 269, 273, 280, 285-286, 309, 322, 341-342
cover story
60, 85, 89-90, 94, 99, 132, 156, 178-180, 182, 207-208, 216, 224, 295
coverage
316
deadstick landing
79
deployment
2, 74, 76, 78, 88-89, 93-94, 134, 161, 165, 167, 247, 250, 255-257, 297, 301-302, 304, 307, 313, 317, 332
design
47
destruction device
177
detachable pogos
47
ejection seat
64, 177, 302
emergency landing
78, 188, 222, 313
film exploitation
115
flameout
71, 78-79, 131, 198, 229
flight plans
82, 177
flights approval process
187-188, 226, 301
foreign pilots
73-74, 321
fuel
5, 23, 47, 61-62, 68, 76, 133, 161, 175, 199, 218-219, 231, 235, 241, 247-248, 271, 274, 277-278, 283-284, 286, 288, 302, 313, 340
fuel consumption
68, 219
fuel gauge
68, 219, 302
fuselage
10, 24, 47, 130, 248, 251, 270, 277, 286
inspection procedures
297
landing gear
10-12, 15, 47-48, 69, 89, 248
low-speed stall
76
maiden flight
66, 288
mission 1104
114
mission 1105
114
mission 1314
117
mission 2003
100
mission 2009
101
mission 2010
101
mission 2013
104-105
mission 2014
105
mission 2020
106
mission 2021
106

mission 2023
108
mission 2029
124
mission 3086
200
mission 3088
200
mission 3089
201
mission 3091
205
mission 3093
205
mission 3095
205
mission 3098
206
mission 3100
206
mission 4016
124
mission 4018
124
mission 4019
126
mission 4030
133
mission 4154
175
mission 4155
168
mission 6011
144
mission length
199
mission plotting
60
mission programmers
78
missions
59, 75, 82, 87, 106, 114, 124, 127, 129, 154, 164, 167, 179, 187, 196, 207, 210, 224, 237, 245, 254, 263, 316-317, 322, 336
Nationalist Chinese crews
222
oleostrut
47, 69
operational altitude
76, 79, 108, 152
Operational Objective system
309
Operational Readiness Inspection
77
outrigger wheels
79
overflights
6, 32, 62, 74, 76-77, 84-86, 88-89, 94-98, 100-101, 104-106, 108-111, 114, 116, 122-128, 135, 138-140, 142-143, 146, 148, 152, 154-157, 159, 161-165, 167-168, 170, 172, 180-181, 195-198, 200-201, 205-206, 208, 210-211, 219, 222-226, 228-237, 239-240, 242-244, 249, 253, 257, 259, 294, 297-301, 310, 315-317, 318-320, 322
paint penalty
149
payload
262, 8, 11, 50, 54, 340, 87, 145, 152
peripheral missions
189, 206, 244, 246
phaseout
257, 310
pilot fatigue
199
pilot hypoxia
80
pilot life-support system
64
pilot selection
60, 283
pilot-survival gear
62
pilot training program
75
pilot urination
64
pilots
42, 299, 47, 56, 59-60,62-66, 68-69, 71-77, 79-80, 94, 109, 113-114, 116, 125, 127, 132-135, 140, 148-149, 153-156, 164, 175-176, 179, 181-183, 186-187, 198-199, 207-212, 214, 221, 223-226, 229-230, 233-236, 238-240, 243, 246-248, 250-253, 256, 283, 319, 321, 332, 340
pilots dessicated
65
pitch-control servo device
296
pogos
47, 79, 108

prebreathing
64, 125
projects
39-40, 42-44, 56, 59, 61-62, 71, 73, 74, 81-82, 88-89, 97, 153-155, 157, 186, 257, 316, 327-331
rocking mount
51, 53
stall envelope
71
stalls
80
tail section
47, 76
taxi trials
68-69
training
302
vertical stabilizers
277, 343
vulnerability study
97, 207
U-2B
95
U-2Cs
152, 196, 208, 210, 217
U-2F
198
U-2G
248, 250, 252, 255
U-2R
251-253, 257, 282, 336
U-3
66
Ukraine
106
Under Secretary of State
188, 225
Unidentified Flying Objects (UFOs)
72
United Arab Republic
153
United Kingdom (See Great Britain)
United Kingdom Requirements Committee
154
United Nations
20
United Press International (UPI)
232
United States
2-4, 7-8, 19-21, 26-27, 73, 77-78, 85, 89-90, 96, 98-100, 108-110, 112, 117, 121, 124, 140, 142, 147-148, 153-156, 159, 161, 163-164, 179, 181-183, 187, 201, 205, 215, 221-222, 224-226, 229-232, 245-247, 254, 256, 285, 294, 297, 299, 301-302, 306, 315, 60, 316, 318-319, 332
unmanned reconnaissance aircraft
291, 304

[redacted] **25X1 and 6, E.O.13526**

UPWIND
303
Urals
22, 163
uranium gaseous-diffusion plant
239
uranium-processing-facility
138
US Air Attache in Moscow
177
US Air Force, Secretary of the
11, 32, 78, 81, 327-329
US Air Force
Air Weather Service
94
Chief of Air Staff
78
EC-130, 146
Inspector General
61
oxygen mask
62
RB-57D Canberra
162
Third Air Force
89
US deterrent force
317
US Geological Survey
254
US House of Representatives
Appropriations Committee
88
US Marine Corps
40, 153, 247, 328
US Marine Corps Reserve
14
US Military Academy at West Point
329
US Navy
Chief of Naval Materiel
285

Neptune bomber
3
Privateer patrol aircraft
3
Naval Ordnance Test Station
340
U.S. News and World Report
99
US nuclear test monitoring system
139
US nuclear tests
147
US Senate, Armed Services Committee
88, 99, 185
US Senate, Foreign Relations Committee
165, 185
US 6th Fleet
153
US 7th Fleet
215
US Intelligence Board (USIB)
196-197, 225, 228, 249, 332, 336
USS Kitty Hawk
247
USS Ranger
250
utility aircraft
66

V

Vance, Cyrus
238, 300, 309-310
Venezuela
211
vernier adjustment
76
very-high-acuity lenses
50
Viet Cong
230, 243
Vietnam
190, 216, 221-222, 230-231, 233-234, 242-243, 246, 301, 304-308, 311, 336, 342-343
Vito, Carmine
65, 105, 124-125, 214
Vladimirovka, Soviet Union
171
Vladivostok, Soviet Union
3, 126
Vogt, Richard
8
Voigt, Woldemar
8
Volga River
22
von Fritsche, Werner
1

W

WADC's Bombardment Branch
8
Wake Island
305
Walker, Christopher H.
155
wallpaper
130
Walter Reed Hospital
96
Warsaw Pact
87
Washington University, St. Louis
29
Watertown Strip
57
Watton RAF Base, England
156
"weather plane" cover story
179
weather research
85-86, 89-90, 132, 182
Weatherwise
216
Wendover, Utah
295
West Potomac Park
42
Western Europe
20, 85, 140
Western Hemisphere
211
"wet-wing" design
68.

WHALE TALE
247, 249
Wheelon, Albert (Bud)
240, 287, 333
Wheelus AFB
121
White House
33, 43, 97, 109, 111-112, 120, 122, 128, 154, 172, 186, 189
Wienberg, Charles F. (Bud)
7, 11-12
Wiesbaden, Germany
80, 95, 104-105, 108, 114-115, 120-121
Wiesner, Jerome B.
37
Williams, John J.
185
Williamsport, Pennsylvania
222
Wilson, Charles
126
wind currents
87
wind tunnels
269
wing teeth
277
wings
5, 8, 11, 23-25, 42, 45, 47, 66, 68, 72, 75-76, 80, 89, 95, 129, 148-149, 199, 242, 248, 251, 271, 274, 277, 286, 312, 340
wiretaps
2
Wonsan Harbor
307
Worcester, Massachusetts
62
world press
85, 122
World War I
1
World War II
1-2, 4, 7-8, 13, 15, 17, 22, 42, 49, 56, 62, 87, 95, 97, 124, 135, 214, 222, 232, 317, 322, 327-331
Wright Air Development Command
4-5, 8, 12, 14, 18, 35, 73
Wright-Patterson
73
WRINGER
2
WS-461L
144-145
Wu-ch'ang, China
228

X

XF-104
(See Lockheed Day Fighter)
X-16
9, 13, 15, 26, 34-35
X-band radar pulses
335
X-rays
291

Y

YAK-25RM
(See Mandrake)
YAK-26
(See Mandrake)
Yakovlev
148
Yakovlev-25
148
Yale University
15
Yancey, William F.
75, 77-78
YC-97 Boeing Stratocruiser
13
Yerevan, Soviet Union
124
YF-12A
285, 291, 294, 308
50z(See also AF-12)
Yokota AFB, Japan
134

Youmen, China
238
Yur'ya, Soviet Union
176
Yutzy, Henry
31

How to Quantify Deterrence and Reduce Critical Infrastructure Risk

Eric F. Taquechel and Ted G. Lewis

ABSTRACT

We propose a definition of critical infrastructure deterrence and develop a methodology to explicitly quantify the deterrent effects of critical infrastructure security strategies. We leverage historical work on analyzing deterrence, game theory and utility theory. Our methodology quantifies deterrence as the extent to which an attacker's expected utility from an infrastructure attack changes after a defender has invested to deter attacks, as compared to their expected utility absent deterrence. We derive expected utilities from a modified game theory approach, which uses probabilistic utility functions, wherein utility function probabilities are functions of investment. We vary these functions based on different information availability assumptions (e.g., perfect vs imperfect attacker information). We produce evidence that it is quantifiably more advantageous to overtly deter, rather than conceal security information, under specific conditions. We also leverage these utility functions to determine the unconditional risk to a defender if deterrence strategies fail, and we determine cost efficiency of those strategies.

INTRODUCTION

DHS policy advocates investing in critical infrastructure (CI) protection to deter terrorist attacks.[1] The *National Infrastructure Protection Plan* is replete with references to deterrence, most notably on page 38: "account for the adversary's ability to recognize the target and the deterrence value of existing security measures." Also, Homeland Security Presidential Directive (HSPD) 7 emphasizes deterrence of CI attacks.[2] Since DHS also advocates quantitative terrorism risk analysis, to support the goal to deter attacks, DHS would benefit from a methodology to quantify the deterrence value of CI terrorism risk reduction investments in a way that complements and enhances existing traditional approaches to risk analysis. Also, DHS would benefit from a working definition of CI deterrence, the lack of which is identified by Morral and Jackson.[3]

Given this policy context, various elements within DHS have begun efforts to analyze deterrence, or influence adversary decision-making before a CI attack is executed. This necessarily involves considering human factors; specifically, thinking about the adversary's approach to terrorism planning. For example, the US Coast Guard, in conjunction with the University of Southern California's Center for Risk and Economic Analysis of Terrorism Events (CREATE), has begun the development of PROTECT, a model intended to help Coast Guard units deter adversary planning by patrolling CI in a random fashion.[4] Recent work at the Naval Postgraduate School in Monterey, CA has produced a model that measures the resiliency of supply chains, accounting for perceived attacker preferences for disrupting a supply chain and increasing commodity shipment costs.[5] Additionally, DHS Science and Technology (S&T) recently undertook a review of multiple methodological approaches to enhancing traditional risk analysis by incorporating insights from intelligent adversary modeling, which accounts for adversary planning and goals.[6]

While well intentioned and promising, these efforts do not explicitly quantify deterrence in the context of critical infrastructure protection (CIP) or explain how to incorporate measurable deterrence into CIP risk equations to show the CI risk reduction effects of deterrence.[7] If DHS components are to implement policies to deter attacks, eventually they may be asked to account to Congress on how they are measuring the effectiveness of that implementation and how that

implementation plays with traditional risk reduction. At present, some DHS agencies such as the US Coast Guard report risk-based performance metrics without explicitly accounting for measurable deterrence. Looking ahead, it is possible that the Coast Guard and other DHS agencies will have to start reporting deterrence effectiveness and resulting risk reduction accounting for this effectiveness.[8] This is our present motivation to develop a quantitative approach to deterrence measurement and to develop a method to incorporate that approach into traditional CIP risk analysis.

In the spirit of this DHS emphasis on deterrence, and assuming DHS components may eventually have to report measures of effectiveness, we propose that CI deterrence can be quantified as the extent to which an attacker's intent to attack a CI target changes. This change is a result of changes in expected utility from a potential attack on a CI after a CI defender attempts to deter an attack, as compared to what the attacker's expected utility from a potential attack would have been before the defender deterred. Utility is the value of an outcome.

Our approach is intended to complement existing, traditional CIP risk analysis tools, not replace them. In fact, our approach relies on input from reputable risk analysis tools. Our example of how we apply this approach focuses on deterring terrorism. However, as long as existing risk analysis tools or best practices can be used to analyze the threat, vulnerability, and consequences of other intentional hazards such as sabotage and theft, our general principles can be applied to measuring deterrence value of measures against any sort of attack on a CI. Indeed, some of our literature review will examine how deterrence principles, which we leverage in our own approach, have been applied to issues other than terrorism. Importantly, our approach does not consider unintentional threats to these CIKR such as natural disasters or industrial accidents[9]. Finally, our approach incorporates elements of the DHS prevent, protect, respond, and recover approach, consistent with Homeland Security Presidential Directive (HSPD) 8.[10]

DETERRENCE AND RISK: MUTUALLY INCLUSIVE IN THEORY

Deterrent tactics include investments to reduce attacker expected utility by influencing attacker capability, target vulnerability, and target consequence. These terms, along with intent to attack, make up the basic DHS risk equation $Risk=T*V*C$. Risk is expected loss to a CI defender from a potential CI attack, but this loss constitutes part of an attacker's gain, or utility. This equation incorporates all aspects of prevention protection, response and recovery as articulated in HSPD-8: reducing threat is preventing an attack, reducing vulnerability is protecting against an attack, and reducing consequences is responding to and recovering from an attack. Thus, if we measure deterrence as a function of changing components of threat, vulnerability and consequence, our deterrence efforts encompass the entire HSPD-8 spectrum.

By influencing these components of attacker utility, a defender may attempt to influence an attacker's belief that the costs of an attack outweigh the benefits or utility, or that the probability of successful attack is too low, so they don't execute that course of action (COA). This is operational deterrence as proposed by Morral and Jackson.[11] Thus, deterrence strategies must convince the attacker than if they execute a certain COA, it will yield benefits less than those yielded by their best course of action without deterrence. Hence, these strategies must account for our attackers' goals as well as our own, as long as we can influence their perspective on attaining their goals. This lends to a game theory approach to quantifying deterrence. If we determine changes in attacker expected utility in a game theory format, we can measure change in intent and thus can measure deterrence.

Deterrence thus inherently includes our adversary's assessment of their risk, or perceived reduction in utility, as well as our assessment of our own risk. Ultimately, when deliberating deterrence strategies, we want to know answers to three questions: (1) what is the extent to which the attacker is deterred, (2) what is the risk reduction, or change in expected defender loss resulting from that

deterrence strategy; and (3) what are the cost implications, or defender deterrence investment efficiency, of that deterrence strategy. All answers must be quantifiable to be useful in quantitative deterrence analysis and terrorism risk management. Taken together, deterrence quantification, defender expected loss as a result of that quantifiable deterrence, and defender investment efficiency constitute a deterrence "portfolio" of criteria a defender must consider. This portfolio would support existing CIKR security plans and risk analysis with additional rigor and measurable deterrence effect data.

GENERAL FORM OF OUR DETERRENCE QUANTIFICATION EQUATION

Formally, quantifiable CI deterrence effectiveness of an lth defender deterrence course of action (COA) given an attacker's kth post-deterrence action, is the difference between attacker pre-deterrence intent, which we denote $Intent_i^{pre}\big|_{\$AB}$, or probability an attacker desires to attack a certain target, and attacker post-deterrence intent $Intent_k^{post}\big|_l$, which is then normalized by the pre-deterrence intent. The i and k subscripts denote specific attacker COA; i is an attacker's pre-deterrence COA and k is an attacker's post-deterrence COA. This is shown in Equation 1.

$$E_l\big|_k = \frac{Intent_i^{pre}\big|_{\$AB} - Intent_k^{post}\big|_l}{Intent_i^{pre}\big|_{\$AB}}$$

Equation 1. Quantification of deterrence

These intent values are functions of attacker expected utilities from an attack. These expected utility values are outputs of a deterrence game in a game theory format.

For the attacker, expected utilities are probabilities of an attack succeeding given it is already desired (attacker capability multiplied by target vulnerability), multiplied by the maximum utility (in dollars) if the attack were to succeed. The attacker's utility or defender's consequence is a combination of deaths[12] and economic consequence (in dollars of annualized profit for that target), and the economic consequence is multiplied by a mitigation factor based on the extent to which the defender could mitigate that economic consequence if an attack was successful.

Threat is generally attacker intent * attacker capability; our methodology only incorporates the capability part of the traditional threat probability equation in the attacker expected utility function. This is because we treat intent as the output of our expected utility calculation methodology, rather than an input, but we apply intent into the defender' s unconditional risk equation to reflect how CI risk changes given quantifiable deterrence so as not to wholly depart from a traditional risk analysis approach. We now elaborate on our approach to leveraging threat with utility functions.

OUR PHILOSOPHY ON THREAT AND UTILITY FUNCTIONS

The risk equation R(risk) = TVC (threat*vulnerability*consequence) is the general standard for CIKR risk analysis because it has a threat component. There are numerous vulnerable and highly consequential CIKR in the US, and leveraging threat judgments helps differentiate and reduce the defender's workload based on the likelihood that an attacker may prefer some target types more than others. However, that threat is often high level and strategic,[13] without accounting for target nuances that may influence attacker perception of specific target attractiveness. Incorporating our quantifiable deterrence, as a function of intent changes based on individual target expected attacker utility, adds this nuance and we believe this more accurately represents unconditional risk.

We claim that intent is not critical component of an attacker expected utility function. We invest to deter and influence desirability or intent. Traditionally, intent is an input to the standard risk equation, as a component of threat. However, we claim that desirability is a function of attacker beliefs about probability of successful attack initiation, execution, consummation, and

consequence of that consummation.[14] Thus, deterrence investments must influence these factors, but desirability or intent is an outcome of that process. Intent then need not be included in probabilistic risk functions or utility functions in a deterrence game.

There are those, such as Louis Cox who believe that threat, or intent and capability to attack, should be an outcome of a probabilistic risk analysis product rather than an input.[15] The outcome of the games in our research will result in intent probabilities. We will show that only intent need be an outcome; we believe that capability probabilities should remain inputs within a risk or utility equation. Intent is still useful in a CIKR risk equation when considering how to allocate resources based on risk, from the defender perspective only. This means if we don't explicitly consider deterrence effects, or changes in attacker intent as a result of changing expected utility, we can justifiably leave intent in the risk equation.

Excluding intent, expected loss should include capability, or the probability an attacker can initiate an attack. The defender invests in offensive, counterterrorism actions to reduce attacker capability to initiate and execute an attack. This is one aspect of investing to deter. If capability reduction investments fail, the attacker is not deterred because they believe a high probability that the attack can be initiated. But, expected loss should also include the probability that the attack can be successfully executed, and will bring about a consequence to the target once consummated, which are components of vulnerability. If vulnerability investments fail, the attacker is not deterred because they may believe a high probability that the attack, once initiated, can be executed and consummated.

Finally, expected loss must include consequence, but only to the extent it is not reduced by defender recoverability and redundancy measures.[16] These measures constitute the final aspect of investing to deter, because if the attacker believes any consequence will be quickly mitigated, they may be deterred as they may not achieve their goals.

Thus, expected utility functions in a deterrence game must include capability probabilities, vulnerability probabilities, maximum consequence values, and probabilities of consequence mitigation. Intent or desirability probabilities are not needed. If intent probabilities were included in expected loss equations, since intent reflects the desirability to carry out an attack, and desirability is a function of probability of successful initiation, execution and consummation, then we would be double counting the deterrent effects of defender investment in those expected utility functions. Thus, our expected utility functions will treat expected utility as a conditional probability: they reflect the expected utility of attack, if the attack already desires to attack, or intent=100%.

DETERRENCE AND RISK: MUTUALLY INCLUSIVE IN PRACTICE

Our utility functions upon which the intent values in Equation 1 are based, account for the traditional risk=threat*vulnerability*consequence approach. In fact, our methodology depends on traditional risk analysis tools to provide this input; it cannot function without existing risk=TVC tools such as the Coast Guard's Maritime Security Risk Analysis Model (MSRAM),[17] TRAM, CARVER, or any legacy probabilistic CI risk analysis tool that leverages threat, vulnerability, consequence, and mitigation data. Thus, our methodology is not intended to supplant existing probabilistic CI risk methodologies and tools, but complement them and extend their functionality. Additionally, myriad CI protection efforts such as Design Basis Threat analysis, Crime Prevention through Environmental Design, and others can easily be leveraged with our proposed methodology, as they contribute to our understanding of threat, vulnerability and consequence-all critical input for our methodological approach. Thus, even if a CIP best practice is not formally considered a risk assessment tool or methodology, if that practice produces data that can meet the basic definition of threat, vulnerability, or consequence, then its input could be leveraged by our deterrence quantification methodology.

Equation 1 will help us answer the first deterrence strategy question. The deterrence game that produces those utilities that support intent values in Equation 1 can also produce defender risk equations, which can be compared to answer the second deterrence strategy question. Further, attacker and defender utilities must be modeled as functions of attacker and defender investment, since return on investment is critical for risk-informed decision making.[18] This will help answer the third deterrence strategy question.

Also, Equation 1 can be applied to quantify deterrence subject to specific theoretical analysis options for how a CI defender and an attacker perceive the utility of different COA. These analysis options include (1) assumptions about information availability circumstances, or the extent to which an attacker knows what a defender is doing to defend their CI; (2) different utility theory assumptions, or assumptions about people assess information and make decisions about the value of an outcome; and (3) assumptions about how risk probabilities are reduced with defender investment.

PRACTICAL IMPLICATIONS AND USEFULNESS OF THIS APPROACH

WHY SHOULD WE HAVE THE ABILITY TO QUANTIFY DETERRENCE?

This quantitative approach can be applied to real-world CI protection. Since we are quantifying deterrence in a way that will leverage traditional probabilistic CI risk data, those involved in CI protection can report a different metric to our policymakers, in a way that reinterprets the effect of CI attack capability, vulnerability and consequence mitigation measures.

To explain, the current performance metric that DHS components report on is how much risk, or expected loss, we have reduced to CI by implementing vulnerability and consequence reduction measures. The focus is on how these measures would defeat an attack already in progress, or how they would mitigate effects after the attack. However, if we leverage this same data to quantify the deterrent value of those measures, we could instead report how we believe we have changed the attacker's expected utility from an attack, and thus have changed their intent to attack, by implementing these measures. Thus, we gain insights into how risk changes based on our adversary's interpretation of our defenses, and we have now accounted for an intelligent adversary. This is the reinterpretation of the effects of our current measures, contextualized for policy that requires both risk analysis and deterrence. Also, current risk methodologies generally only account for an engineering approach to risk management, as they focus primarily on the details of vulnerability and consequence. They include threat as well, but often this is a very high-level, non specific threat probability estimation. So, our methodology will account for more tactical aspects of threat probabilities, such as how individual target engineering characteristics such as security measures actually influence the attacker's interest in attacking, and thus threat judgments can be more granular and target-specific.

That said, it seems necessary to be able to incorporate our quantitative deterrence effectiveness results back into the standard risk=TVC equation. There are theoretical challenges with this, but we believe they are easily overcome. In brief, adopting our deterrence quantification methodology and incorporating its results in back into traditional probabilistic risk=TVC, where T represents intent and capability, might seem to double count the effects of our capability, vulnerability, and consequence factor reduction deterrent measures. It would count those deterrence effects once against those factors themselves, and again in intent. However, we claim that unconditional risk, or risk including the probability the attacker desired to attack after deterrence measures are implemented, appropriately double-counts the effects of those deterrence investments: once in the reduction of capability, vulnerability, and/or consequence, and again in intent. We must account for the likelihood the attacker wants to attack, and the likelihood of success if they do initiate an attack. The current form of risk used for CIP is unconditional risk where we believe the

intent probability modifies conditional risk, and conditional risk is capability*vulnerability*consequence: the expected loss given an attack is already desired with 100% intent.

This approach is useful for a wide audience: academics who will consider the technical aspects of our methodology, CIP practitioners who may implement the deterrence measures that inform our methodology once the methodology is accepted within academia; and DHS agencies who must report performance effectiveness measures to Congress and other oversight bodies after practitioners have implemented the measures.

We expound on how we incorporate measurable deterrence into the traditional risk equation in a case study example. We also give examples of how to apply this methodology to notional real world CI in a subsequent section of this research. Finally, we give examples of how quantifiable deterrence effectiveness might change given different information availability circumstances, as that seems the most practical of the theoretical analysis options we have introduced: a practitioner naturally wants to know whether they should make their security measures overt or covert to deter! The other two analysis options are areas of interest but we do not give examples of how deterrence effectiveness might change if those options are varied; we save this for future research. We discuss the theories behind those analysis options briefly in our analysis of the need for these options and our literature review.

WHY THE THEORETICAL ANALYSIS OPTIONS? DIFFERENT DETERRENCE EFFECTIVENESS DATA MAY RESULT

Equation 1 can be applied to quantify deterrence subject to specific theoretical analysis options for how a CI defender and an attacker perceive the utility of different COA. These analysis options include assumptions about information availability circumstances, or the extent to which an attacker knows what a defender is doing to defend their CI. In game theory modeling, whether actors have perfect (known) or imperfect (unknown) information about their opponents is critical to understanding the game's outcomes. Since we will be leveraging principles from game theory in our approach, we focus on this analysis option. These theoretical analysis options also include different utility theory assumptions, or how assumptions about people assess information and make decisions about the value of an outcome. Nikhil Dighe et. al. urge that future work on analyzing deterrence investment should consider alternatives to traditional rationality, or subjective expected utility (SEU).[19] Utility functions can assume either (SEU) or an alternative, prospect utility theory (PU).

Finally, these analysis options include assumptions about how risk is reduced with defender investment. In theory, probabilities of attack success may change linearly with investment, but in reality, the relationship may be an exponential function of investment; it becomes increasingly difficult to reduce risk as more risk is reduced. Dighe et. al. advocate exploration of the effects of security investments when those investments do not reduce the success probability of attacks to zero.[20] This nonlinearity can be observed in engineering problems; for CI security measures, both engineering factors and social factors such as public resistance to increased security may contribute to nonlinearity.

The practical implications of experimenting with the different theoretical analysis options are numerous. Uncertainties facing an attacker are critical to understanding deterrence, but are rarely leveraged in game-theoretic analysis of counterterrorism.[21] We have done initial work on this and have produced evidence suggesting whether it is more advantageous for a CI defender to make defensive investments unknown, or imperfect, to would-be attackers rather than perfectly known. The defender advantage in such cases would be lower average unconditional risk after deterrence under specific information availability circumstances.

If evidence supports an advantage in deterrence effectiveness under imperfect information circumstances, implications are that a government should consider classifying its CI defensive investments and any analysis

of how those investments reduce infrastructure vulnerability and consequence. Conversely, if evidence supports an advantage of perfect information, a government should instead actively communicate to would-be CI attackers not only its defensive investments but also detailed calculations of how those investments reduce vulnerability and consequence. This approach assumes the attacker believes the defender's disclosure of investment and effectiveness information is accurate.

In order to generalize these findings, any advantage of a specific information availability circumstance must be robust given different utility theory assumptions, and also given different probability-investment relationship assumptions. Behavioral science and historical research support the need to analyze decision making under nonlinear or biased circumstances and thus examine different assumptions about utility when we analyze deterrence. Nonlinear engineering phenomena support the need to analyze how system performance, or CI attacker capability, vulnerability, and consequence mitigation in CIP research, changes nonlinearly as a function of effort or investment expended. We do not give examples of sustained defender advantage from a specific information availability circumstance across these different analysis option assumptions in this paper, but this is an area for future research.

DISCUSSION OF RELEVANT WORK

We now examine relevant work to create context for our general methodology to quantify CI deterrence, and for our specific theoretical analysis options. We integrate different aspects of this work with our own techniques into a methodology to quantify CI deterrence and create a deterrence portfolio. There is historical work to measure deterrence effects in other areas, including fishery enforcement, drug smuggling, and aviation terrorism, but DHS currently lacks a methodology to explicitly quantify CI terrorism deterrence. However, there is plenty of research on quantitative risk measurement, analysis, and management, which suggests we should leverage quantitative risk to quantify deterrence.

RISK, RETURN ON INVESTMENT, AND DETERRENCE

DHS defines risk as a function of threat, vulnerability, and consequence.[22] For terrorism, threat is defined as the likelihood of an attack being initiated, accounting for both adversary intent and capability. Intent is the probability an attack is desired, and capability is the probability it can be executed. Vulnerability is the likelihood an attack is successful given it is attempted. Finally, consequences are the effects a terrorist attack. Thus, this approach to analyzing terrorism risk is probabilistic in nature. The probabilistic approach to risk analysis has been widely accepted for some time.[23]

Since DHS is using probabilistic risk analysis techniques for counterterrorism,[24] the relationship between probabilistic risk analysis and general deterrence theory must be examined before we attempt to quantify deterrence effectiveness of CI protection strategies. In general, there is much precedent for integrating risk analysis with deterrence. Richard Lebow and Janet Stein propose that deterrence effectiveness depends upon an actor's attitude toward risk.[25] Elaine Bunn advocates that deterrence policies must account for adversary attitudes toward risk.[26] However, these arguments are not specifically focused on CI attack deterrence, and contain no explicit quantitative calculations of probabilistic risk that would support deterrence quantification.

Other literature focuses on possible relationships between CI risk analysis and deterrence. For example, William McGill and Bilal Ayyub claim that deterrence is a function of infrastructure vulnerability, a component of infrastructure risk.[27] In fact, the probability an attack would be initiated, or threat, is claimed to be a function of the perceived target attractiveness, and the target attractiveness is a function of the perceived probability of success, or vulnerability. Thus, a lower vulnerability target means a less attractive target; a less attractive target then

means a lower probability of attack initiation; and a lower attack initiation probability then means the attacker has been deterred. This suggests a fundamental relationship between probabilistic risk and deterrence. We must now examine deterrence theory more closely and put our proposed definition of CI deterrence measurement in context of this historical discussion.

DETERRENCE THEORY

Deterrence theory has been vigorously discussed and debated for decades. Patrick and Audrey Cronin propose that deterrence occurs when an actor discourages aggression toward another actor, with the intended outcome that the former never has to respond to aggressive action by the latter.[28] In other words, the goal of deterrence is to convince an adversary to refrain from aggressive action. Lebow claims that deterring involves manipulating an adversary's assessment of a situation by convincing them that costs of acting exceed benefits.[29] Finally, putting theory into practical guidance the Department of Defense Deterrence Operations Joint Operating Concept specifies that deterrence affects adversary decision making in three ways: (1) imposing costs on an undesirable adversary course of action; (2) denying the perceived benefits of such a COA; and (3) encouraging restraint by making it seem more attractive than action.[30]

These approaches to deterrence suggest that the only way to deter is to convince an adversary not to attack. But, applying this approach to deterring terrorist attacks on CI is problematic for two reasons. One, the United States has so many CI that spending resources to prevent attacks on all seems inefficient. Two, these deterrence theories were developed during the Cold War when nation-states could be deterred by threat of mutually assured destruction. With asymmetrical warfare and terrorism, convincing our adversaries we can eliminate them is a challenge since they are difficult to find. Thus, a modified definition of deterrence is needed for CI protection. An all-or-nothing approach applied to DOD tactic (1) is impractical because it is difficult to impose costs. For (2), denying benefits would mean defending all infrastructures, which we attempt to do at great expense. Finally, for (3), our adversaries have stated their intent to continue to attempt to inflict damage; restraint will likely never be acceptable to organizations such as Al-Qaeda whose raison d'être is to inflict damage upon the United States and its allies. Thus, our form of deterrence for CI protection – influencing an adversary's decision making process such that their expected utility from attacking a CI changes after we deter by investing – is appropriate.

The idea of accepting losses is not without precedent. For example, John Major proposes the concept of equilibrium expected loss (EEL): a defender moves resources to various high-risk infrastructures, making them equally unattractive to an attacker and creating an EEL. Any unprotected infrastructure, if attacked, will yield a gain to the attacker that is less than the EEL.[31]

Two critical components of deterrence are credibility and signaling. Bunn claims that credibility is critical: to deter, we must be able to signal to the adversary we have both the credible capability and the will to impose costs or deny benefits.[32] Thus, for CI deterrence, a terrorist must believe that we can deny them benefits by defending our infrastructure. That belief can be reinforced by the terrorist's own observations of our defenses, and our explicit signaling of those defenses. The former type of reinforcement begs the question of whether the terrorist accurately perceives our defenses, and the latter begs the question of whether the terrorist believes our disclosures. These issues have been researched; for example Erik Jenelius et. al. analyze the effects of adversary perception on their decision making.[33] The present research assumes that prior to deterrent investments at the start of a deterrence game, all information about the defender's original "pre-deterrence" defenses is known to an attacker, and subsequent deterrence game outcomes determine how deterrence is quantified depending on whether the attacker has perfect or imperfect information about the defender's deterrence COA. Thus, our signaling of information should be a key determinant of how quantifiably effective our deterrence investments are. There is the possibility that

an attacker could infer weaknesses if we signal prospective defenses before they are implemented. Importantly, we do not advocate actually signaling the existence of security or response measures before they are implemented. We only argue that when we are considering what deterrence investments to implement, we should determine whether making those investments transparent or obscured, once actually implemented, would influence the attacker more. Also, we can run our deterrence game simulation with as many CIKR targets as desired, using computer programs if necessary, to account for the possibility that signaling defenses at targets will focus the attacker attention elsewhere. We now know there is a relationship between probabilistic risk analysis and deterrence, and we have a working definition of CI attack deterrence as a change in attacker intent. Thus, we will use a probabilistic risk approach to quantify deterrence. Next, we examine previous and ongoing efforts to analyze and quantify deterrence.

ANALYZING DETERRENCE

There is a considerable repository of literature offering insights into how deterrence might be quantified, without ever explicitly stating how to quantify it for CI protection. In general, the literature shows that analysis of drug smuggling, fisheries law enforcement, critical infrastructure security, and aviation security, reveals that lowering the chances of successful attack or violation of a law, and communicating the imposition of penalties if caught in violation, tended to deter would-be violators. One study gleaned that the subjects interviewed would be deterred if they knew the high probabilities of failure if they attempted to violate a law, but they had difficulty estimating those probabilities with any rigor. Other studies more explicitly focused on critical infrastructure security, postulating that the change in expected utility or outcome of a CI attack would influence the probability an attacker would want to attack in the first place. Much of the more recent literature leverages ideas from utility theory and game theory, and thus we leverage the ideas of influencing probabilities and consequences, along with game theory and utility theory, in our approach to quantify deterrence. We now expound on these theories.

GAME THEORY AND THEORETICAL ANALYSIS OPTIONS

Game theory is an analytical approach designed to help understand interactions between competing or cooperating decision makers. It treats decision makers as "players" participating in a "game" with certain rules and assumptions.[34] Two of these assumptions are that decision makers are rational and that they simultaneously account for their opponents' anticipated behavior when choosing a course of action. Game theory is an appropriate tool for studying the strategic interaction between governments and terrorists.[35]

Game theory assigns an expected "payoff" value or expected utility from certain courses of action to each player. A payoff is the subjective value of the outcome of a course of action. It is assumed that each player wants to maximize his own payoff or utility.[36] But, game theory often predicts an outcome in which each player has a payoff which may be less than that which they could gain without the influence of their opponent. Our approach will create payoffs or utility functions to be used in a game, from traditional risk components: capability, vulnerability, and consequence. Ultimately, utility is a function of probabilistic risk. We treat the defender's probabilistic loss as the attacker's gain or utility in the present research. We also derive detailed expected utility functions for an attacker, reflecting the belief that terrorists perform cost-benefit analyses.[37]

We also examine how perfection or imperfection of attacker information, assumptions which we call information availability circumstances, influences the quantifiable effectiveness of deterrence and other components of our deterrence portfolio. In traditional game theoretical approaches, information imperfection, or lack of information about all opponent moves, lends to a pure strategy Nash Equilibrium (NE) as the predicted outcome

of the game, whereas perfect information generally lends to a mixed strategy NE as the predicted game outcome in a Stackelburg leader-follower game.

However, we do not follow the traditional game theoretical approach in our research. In essence, we eschew the traditional game theoretical approach of leveraging a pure Nash Equilibrium (NE) or a mixed strategy, to account for the likelihood that adversary decision making is more complicated that what a simple equilibrium predicts. So, for distinguishing between the effects of information availability upon deterrence effectiveness, we create utility functions in our deterrence game that reflect either perfect attacker information, or one of what we propose are three possible "information imperfection biases" that the attacker is assumed to have: neutral, conservative, and overconfident. These biases are part of assessing the adversary in a design-basis threat approach.

A neutral bias means an "average attacker", who will assume the defender deters by reducing target consequence and vulnerability by 50% of what they were prior to deterrence investment. A conservatively biased attacker will assume the defender deters by reducing target consequence and vulnerability by 95% of what they were prior to deterrence investments,[38] and an overconfidently biased attacker will assume the defender invests nothing to deter; thus post-deterrence attacker expected utilities are the same as pre-deterrence expected utilities. We offer these biases as a start to exploring what we call a "theoretical analysis option" for quantifying deterrence; more research into modeling attacker decision making under imperfect information conditions is an area for future research. There is some literature on this subject, for example see Jenelius et. al.[39]

There are additional "theoretical analysis options" to apply to our methodology. These options will allow us to vary the composition of the utility functions that we will use in our methodology to quantify deterrence. Game theory leverages utility functions and we use a basic game in our approach, and so we expound upon utility theory and explain different views on how people assess subjective value in our extended literature review. In brief, subjective expected utility claims people make decisions linearly; whereas prospect utility claims people make decisions nonlinearly, subject to biases. There is experimental evidence to support nonlinear decision-making; hence a well-informed approach to quantifying deterrence must consider these different ideas as a sensitivity analysis. Prospect utility would require us to modify components of our utility functions with weights and other modifiers to reflect nonlinear biased decision-making.

The other theoretical analysis option is probability-investment relationships. We model utility function components such as capability and vulnerability probabilities as functions of investments, to capture cost information for return on investment calculations. That said, evidence from engineering sciences supports a nonlinear relationship between effort applied to solve a problem and the extent to which the problem is solved. There are linear and nonlinear (exponential) ways to model how these probabilities change as we invest more money to reduce them.

Now that we have examined literature on how deterrence might be quantified, and introduced theories that are relevant to creating the utility functions in a game theoretical approach to quantifying deterrence, we will explain our general methodology to quantify deterrence in context of a case study. Also, we will show how to quantify defender risk as result of the deterrence strategies in the case study, and we will show how to determine deterrence investment efficiency or ROI. Before we move to our case study, we summarize the applicability of our different theoretical analysis options to CI attack deterrence in Table 1:

Information Availability Circumstance	Perfect Information	Applied if we believe the attacker knows everything about the defender's deterrence investments
	Imperfect Information	Applied if we believe the attacker does not know everything about the defender's deterrence investments; the attacker biases their estimates based on whether they are overconfident, conservative, or neutral decision makers
Utility Theory	Subjective Expected Utility	Applied if we believe the attacker analyzes probabilities of attack success and probabilities of producing consequences, in a linear fashion
	Prospect Utility	Applied if we believe the attacker analyzes probabilities of attack success and probabilities of producing consequences in a nonlinear fashion. They bias toward overweighing expected utility when in the domain of loss, or behind a reference point. Thus, an attacker may be more likely to prefer attacking a target that objectively is very hard to attack or would suffer minimal consequence, if a successful attack will put them closer to their goals or reference point.
Probability-Investment Relationship	Linear	Applied if we believe the attacker's and defender's investments change attack success and consequence mitigation probabilities linearly (e.g. it is possible to attain 100% vulnerability elimination)
	Exponential	Applied if we believe the attacker's and defender's investments change attack success and consequence mitigation probabilities nonlinearly, or more specifically exponentially (e.g. it is impossible to attain 100% vulnerability elimination; and it becomes progressively harder to reduce the same amount of vulnerability with the same dollar amount, as we continue to reduce vulnerability)

Table 1: Applicability of Theoretical Analysis Options to CIP

DETERRENCE QUANTIFICATION METHODOLOGY - A REAL WORLD CASE STUDY

Since we are proposing a methodology to support risk analysis with deterrence measurement, and risk analysis is largely a quantitative process, we develop a technical approach to quantify deterrence. We summarize the results of this technical approach here in the context of a case study, focused on the potential deterrence effectiveness of notional FEMA Port Security Grant Program (PSGP) infrastructure security grants.[40] In general, we must compare attacker expected utilities from both before and after deterrence investments, those investments being PSGP grants improve CIKR security, in order to quantify deterrence.

In our example we create deterrence portfolios showing the quantification of

deterrence given certain deterrence COA, how defender unconditional risk changes as a result of that quantifiable deterrence, and defender investment efficiency or ROI of those deterrence COA. We also can show how the results of these portfolios vary based on information availability circumstances: either the attacker has perfect information and defender deterrence COA perform a certain way, or the attacker has imperfect information and the defender deterrence effectiveness may vary depending on the attacker's information imperfection biases. We assume subjective expected utility and a linear probability-investment relationship throughout our example in this paper; future work would analyze deterrence effectiveness under different utility theory and probability-investment relationship assumptions.

In our example we leverage notional CI risk data from an existing DHS CI risk analysis tool, the US Coast Guard's Maritime Security Risk Analysis Model (MSRAM), though any CI risk analysis tool that leverages threat, vulnerability, consequence data, and mitigation measures could be used. Our methodology incorporates MSRAM risk data and extends the interpretation of that data. We will demonstrate how our methodology can be applied across multiple DHS CI sectors, as our notional CIKR competing for grant allocations to deter attacks are in different CIKR sectors[41].

Importantly, in our example we assume the federal government allocates the deterrence funding. Obviously, this assumption can be changed depending on the scenario modeled. We also assume an attacker is not motivated to attack a specific target type (e.g., an eco-terrorist only wanting to attack chemical refineries). Rather, the attacker is expected to want to maximize its possible utility, irrespective of target type. But, the methodology could be applied to include only targets of a specific type in future research.

CASE STUDY: QUANTIFYING DETERRENCE EFFECTIVENESS OF DEFENDER INVESTMENTS, AND RESULTING TERRORIST ATTACK RISK REDUCTION FOR A CHEMICAL FACILITY AND A FERRY TERMINAL

Suppose we want to determine the quantifiable deterrence effectiveness of different CI investments at two different targets: a maritime chemical facility which we denote target A, and a ferry terminal which we denote target B. The "defender" is the government and target owner/operators working in collaboration, and the design-basis threat for the "attacker" is a non-specific terrorist organization that attacks these targets in one specific way (e.g., boat bomb). Our two notional targets are in different DHS CI sectors; yet they often exist together in a local operational environment such as a port. We note that we only use two targets for simplicity in illustrating the methodology; however our methodology can be applied to as many targets as desired. The deterrence game can be expanded as necessary; computer programming may be required to capture CIKR in a large game.[42]

The defender has already invested to reduce vulnerability and consequence at these targets, and thus we can estimate how we have reduced risk in the traditional sense. However, given DHS encouragement to examine deterrence, and given that agencies currently report risk reduction performance effectiveness metrics to Congress and thus conceivably may be required to report deterrence effectiveness and resulting risk reduction in the future, we now want to analyze the potential deterrent effects of hypothetical future deterrence investments, in a quantifiable way.

There are eleven basic steps to quantifying deterrence and creating our deterrence portfolio to address all three deterrence strategy questions introduced at the beginning of this paper. In sum, there are three phases of analysis: pre-deterrence analysis, post-deterrence analysis, and comparison to quantify deterrence and create the deterrence portfolio. This portfolio would add data to the existing library of CIKR security plans and supporting decision-

making. We discuss these three analysis phases in context of our case study.

PRE-DETERRENCE ANALYSIS FINDINGS

Pre-Deterrence Attacker Expected Utilities

The notional expected utility values in Table 2 include capability to attack, vulnerability to attack as a function of security investments, and consequence mitigated by recovery/ response measures. For example, suppose the intelligence community estimates a terrorist organization's capability to initiate an attack with a boat bomb against either type of facility. Also, prior to implementation of PSGP grants intended to deter, suppose both the chemical facility and ferry terminal have measures to detect an attack in progress (e.g., cameras); training and equipment to engage an attack before it reaches the focal point of the infrastructure for detonation (such as armed guards stationed within a reasonable response time from where an attack would initiate); and has the means to defeat that attack (guards have proper weapons and training). Finally, suppose each facility has business continuity plans and coordinates with stakeholders to ensure recoverability if attacked. These are all pre-deterrence mitigation capabilities. These data are assumed to represent (1) the attacker's perceptions of their attack capabilities and (2) target vulnerability and consequence based on attacker "scoping" reflecting extensive research and observations of the above-listed target attributes.

Attacker *i*th attack option	**Notional expected utility value ($)** $U_e T_i^{pre}$
Attack chemical facility	$257,142.86
Attack ferry terminal	$763,636.36
Attack both	$1,020,779.22
Refrain from attack	$0

Table 2: Notional attacker pre-deterrence expected utilities

Table 2 shows that given our notional pre-deterrence vulnerability, and consequence data for the chemical facility and ferry terminal, and attacker capabilities to attack each respectively, attacking both targets simultaneously yields the greatest expected utility to the attacker. This is intuitive, but we quantify that intuition here in dollars, and in comparison to the utility of other attack options.

Pre-Deterrence Attacker Intent Values

Given the above expected utilities, we can create intent values:

Attacker *i*th attack option	Notional intent value (%) $Intent_i^{pre}\Big\|_{\$AB} = \frac{U_e T_i^{pre}\Big\|_{\$AB}}{\sum_{i=4}(U_e T_i^{pre}\Big\|_{\$AB})}$
Attack chemical facility	12.60%
Attack ferry terminal	37.40%
Attack both	50.00%
Refrain from attack	0.00%

Table 3: Notional attacker pre-deterrence intent values

Table 3 shows that attacker intent to attack both targets simultaneously is greatest. Since Table 2 showed the greatest expected utility to results from this attacker option, it is no surprise that this option results in the greatest attacker intent value.

Pre-Deterrence Defender Unconditional Risk

We then apply these intent values in Table 3 to conditional risk values, to get unconditional risk values:

Attacker *i*th attack option	Notional unconditional risk value ($) $Risk_i^{pre}\Big\|_{\$AB} = \frac{(Cap_i^{pre} V_i^{pre} Con_i^{pre}\Big\|_{\$AB})^2}{\sum_{i=4}(U_e T_i^{pre}\Big\|_{\$AB})}$
Attack chemical facility	$32,388.22
Attack ferry terminal	$285,634.98
Attack both	$510,389.61
Refrain from attack	$0.00

Table 4: Notional defender pre-deterrence unconditional risk values

Not surprisingly, Table 4 shows that the unconditional risk of having both targets attacked simultaneously is greatest. So, going forward we can expect that the attacker's greatest expected utility before we deter would result from attacking both targets simultaneously; thus their intent is greatest. And, our unconditional expected loss, or risk, from having both targets attacked simultaneously, is our greatest risk.

POST-DETERRENCE ANALYSIS FINDINGS

Post-Deterrence Attacker Expected Utilities

We note that there are sixteen attacker expected utilities post-deterrence, as opposed to four pre-deterrence, because the defender has four different deterrence investment options. These options are (1) to invest PSGP

grant money at only the chemical facility (for instance to increase the CCTV capabilities to detect); (2) invest grant money at only the ferry terminal (for instance to provide more training and equipment for terminal security personnel); (3) award grants to both; or (4) award grants to neither.[43] The attacker post-deterrence expected utilities are as a function of these investment options are:

Attacker *k*th attack option	**Notional expected utility value ($), defender invests at chemical facility** $U_e T_k^{post}\big\|_{\$A}$	**Notional expected utility value ($), defender invests at ferry terminal** $U_e T_k^{post}\big\|_{\$B}$	**Notional expected utility value ($), defender invests at both** $U_e T_k^{post}\big\|_{\$AB}$	**Notional expected utility value ($), defender refrains from investment** $U_e T_k^{post}\big\|_{\$0}$
Attack chemical facility	$57,142.88	$257,142.86	$57,142.86	$257,142.86
Attack ferry terminal	$763,636.36	$190,000.00	$190,000.00	$763,636.36
Attack both	$820,779.22	$447,142.86	$247,142.86	$1,020,779.22
Refrain from attack	$0.00	$0.00	$0.00	$0.00

Table 5: Notional attacker post-deterrence expected utilities, defender deters by investing in four different ways

Table 5 shows that given our notional post-deterrence vulnerability, and consequence data for the chemical facility and ferry terminal, and attacker post-deterrence capabilities to attack each respectively, attacking both targets simultaneously yields the greatest expected utility to the attacker, for each of the four defender deterrence COA. This notional data may reflect defender efforts to reduce capability (prevent), reduce vulnerability (protect), and reduce consequence (response and recover), though in this example the specific PSGP grants are intended to reduce vulnerability. Overall the attacker would gain the greatest expected utility if they attacked both and the defender invested nothing in deterrence, as seen in Table 5. This seems intuitive, but here we quantify this intuition.

Post-Deterrence Intent Values

Given the expected utilities described above, we can create sixteen intent values.

Attacker *k*th attack option	Notional intent value (%), defender invests at chemical facility $Intent_k^{post}\big\|_{\$A}$	Notional intent value (%), defender invests at ferry terminal $Intent_k^{post}\big\|_{\$B}$	Notional intent value (%), defender invests at both $Intent_k^{post}\big\|_{\$AB}$	Notional intent value (%), defender refrains from investment $Intent_k^{post}\big\|_{\$0}$
Attack chemical facility	3.48%	28.75%	11.56%	13.62%
Attack ferry terminal	46.52%	21.25%	38.44%	37.55%
Attack both	50.00%	50.00%	50.00%	50.00%
Refrain from attack	0.00%	0.00%	0.00%	0.00%

Table 6: Notional attacker post-deterrence intent values, defender deters by investing in four different ways

Table 6 shows that attacker intent to attack both targets simultaneously always dominates, as would be expected since attacker expected utility is always greatest.

Post-Deterrence Defender Unconditional Risk

We then apply these intent values in Table 6 to conditional risk values, to get unconditional risk values.

Attacker *k*th attack option	Notional unconditional risk value ($), defender invests at chem. facility $Risk_k^{post}\big\|_{\$A}$	Notional unconditional risk value ($), defender invests at ferry terminal $Risk_k^{post}\big\|_{\$B}$	Notional unconditional risk value ($), defender invests at both $Risk_k^{post}\big\|_{\$AB}$	Notional unconditional risk value ($), defender invests from investment $Risk_k^{post}\big\|_{\$0}$
Attack chemical facility	$1,989.15	$73,938.84	$6,606.11	$32,388.22
Attack ferry terminal	$355,235.90	$40,367.41	$73,034.68	$285,634.98
Attack both	$410,389.61	$223,571.43	$123,571.43	$510,389.61
Refrain from attack	$0.00	$0.00	$0.00	$0.00

Table 7: Notional defender post-deterrence unconditional risk values, defender deters by investing in four different ways

Table 7 shows that regardless of what the defender does to deter, the greatest risk is of having both targets simultaneously attacked; but intuitively – by investing at both targets – this risk is minimized. These unconditional risk scores reflect the vulnerability reduction effects of the notional grant investments, but also reflect the deterrence effects (changing attacker intent) of the investments. Now, we compare this pre-deterrence and post-deterrence data to develop our deterrence portfolio.

COMPARING PRE-DETERRENCE AND POST-DETERRENCE DATA

Quantification of Deterrence

We compare intent values from Tables 3 and 6, using Equation 1 to quantify the deterrence effectiveness of each defender COA, given all possible attacker options.

Attacker *k*th attack option	Deterrence Effectiveness (%), defender invests at chem. facility $E_{\$A}\big\|_k$	Deterrence Effectiveness (%), defender invests at ferry terminal $E_{\$B}\big\|_k$	Deterrence Effectiveness (%), defender invests at both $E_{\$AB}\big\|_k$	Deterrence Effectiveness (%), defender invests from investment $E_{\$0}\big\|_k$
Attack chemical facility	72.36%	-128.29%	8.22%	0.00%
Attack ferry terminal	-24.37%	43.23%	-2.77%	0.00%
Attack both	0.00%	0.00%	0.00%	0.00%
Refrain from attack	N/A	N/A	N/A	0.00%

Table 8: Notional quantification of deterrence effectiveness of four different defender deterrence COA

Notice that investing at target A reduced attacker intent to attack target A nearly 72%, whereas the negative sign for attacking B means the attacker is actually incentivized to attack target B. Thus, this deterrence COA is quantified as 72% effective at deterring an attack against target A, whereas it is 24% *ineffective* at deterring an attack against target B. It is intuitive that we would transfer intent to B by deterring attacks on A, and we can quantify this intuition here. Now, we want to know the deterrence effectiveness of investing at target B.

As expected, investing at target B incentivizes the attacker to attack target A, by increasing intent 128%, whereas it reduces intent to attack B and thus deters by 43%. Thus, this transfers intent to attack target A. Given that we have to make a decision on where to invest to deter, would we rather invest at A or B? To support this decision, we must know how defender unconditional risk changes given each possible deterrence COA. Transfer of intent and quantifiable deterrence effectiveness are useful metrics but are also means to an end: determining the resulting change in unconditional risk.

Finally, we examine the deterrence effectiveness of investing at both targets. The attacker is marginally incentivized to attack target B, perhaps because of the large consequence as compared to that of target A,

and the attacker is slightly deterred from attacking target A in response. As with the previous two deterrence investment options, we must determine the resulting defender conditional risk across all attacker options, if the defender deters by investing at both targets. We note that the deterrence effectiveness of refraining from investment is 0% for all attacker options; attacker post-deterrence intent does not change from pre-deterrence intent if the defender does nothing.

New Defender Unconditional Risk

Given this deterrence quantification data, we now analyze the second component of a deterrence portfolio, or how defender unconditional risk changes as a result of quantifiable deterrence. We leverage Table 7 data and aggregate defender unconditional risk given an *l*th deterrence COA.

Deterrence COA	Aggregate defender risk ($) $\sum_{k=4}(Risk_k^{post}\big\|_{\$l}) = Risk_A^{post}\big\|_{\$l} + Risk_B^{post}\big\|_{\$l} + Risk_{AB}^{post}\big\|_{\$l}$
Invest at chemical facility	$767,704.66
Invest at ferry terminal	$337,877.67
Invest at both	$203,212.22
Invest at neither	$828,412.81

Table 9: Comparing aggregate defender unconditional risk given certain defender deterrence investments

Overall, the best quantifiably effective defender deterrence COA is to deter by investing at both the chemical facility and the ferry terminal, as the resulting unconditional risk is lowest. The least quantifiably effective deterrence COA is to do nothing, as the risk is highest. From a deterrence effectiveness and unconditional risk reduction perspective, these seem intuitive decisions, but with our methodology we can quantify the deterrent effect of these COA and quantify the impacts on risk reduction, relative to the impacts of other COA. Thus we see relative strengths and weakness of each option, and a decision maker can improve their risk reduction and deterrence performance measurement reporting beyond just reporting "we've invested to deter and reduce risk." Also, this technique of aggregate defender unconditional risk across all attacker options assuages the concern with transferring intent or risk from one target to another;[44] if we simply compare all possible outcomes before deterrence and after deterrence, we now have a metric that acknowledges possible transfer of intent between targets, but looks at the big picture. We believe a "transfer of intent" does not necessarily equate to a transfer of risk. We must consider the capability to execute an attack, vulnerability to attack, and consequence of an attack on all targets in the game. If the resulting intent to attack target B, when combined with these other risk factors, renders an overall lower unconditional risk to B than the unconditional risk of an attack on target A, in fact we have not transferred risk.

Return on Investment

Suppose the grant to defend the chemical facility with additional CCTV is for $1 million; the grant to defend the ferry terminal by training and equipping security personnel costs $2 million, and $3 million is the cost to defend both simultaneously. We show the ROI of each deterrence COA:

Deterrence COA	ROI (unitless) $ROI_l = \frac{\frac{Risk_{AB}^{pre}\big\|_{\$AB} - \frac{\sum_{k=4}(Risk_k^{post}\big\|_l)}{4}}{Risk_{AB}^{pre}\big\|_{\$AB}}}{\$_l}$
Invest at target A	2.66*10^-7
Invest at target B	3.01*10^-7
Invest at both	3.12*10^-7
Invest at neither	N/A

Table 10: ROI of certain defender deterrence investments (ROI unit-less since risk reduction in $)

Notice that there is no ROI of refraining from investment, as we cannot divide by $0. Since we invested the same total dollar amount for each *l*th deterrence COA in this example, we get the best ROI from investing in both targets, as that created the lowest aggregate defender unconditional risk as seen in Table 9. This is an additional performance measurement for a decision maker to report. However, if we were to use different dollar values for the different *l*th COA, the ROI rankings might be different.

Using Tables 2 through 10, we can now create deterrence portfolios for each *l*th deterrence COA. For example the deterrence portfolio for *l*th COA=investing at target A, with respect all possible *k*th attacker post-deterrence options, is:

$$\left[\begin{array}{c} \begin{bmatrix} E_{\$A}\big|_A = 72.36\% & E_{\$A}\big|_B = -24.37\% \\ E_{\$A}\big|_{AB} = 0\% & E_{\$A}\big|_0 = n/a \end{bmatrix} \\ \overline{Risk_k^{post}}\Big|_{\$A} = \$191{,}903.67 \\ ROI_{\$A} = 2.66*10^{-7} \end{array}\right]$$

Figure 1: Values for deterrence portfolio, defender deters by investing at chemical facility alone (target A), perfect information

In Figure 1, $\overline{Risk_k^{post}}\Big|_{\$A}$ represents average unconditional risk resulting from deterring at target A. We can repeat this consolidation of Tables 2 through 10 for the remaining three deterrence COA, which will show that investing at both targets is the best option for reducing risk and ROI. These deterrence portfolios are a succinct way for government agencies and CI owners/operators to report the deterrence effectiveness of deterrence

measures, and resulting CI risk reduction and ROI. Decision makers may value different components of these portfolios differently; if both the government and CI operators will invest to deter, then they must decide what the most important component of the deterrence portfolios are. If industry is the only contributor of deterrence funding, then perhaps ROI may be the most important component. If government is the only contributor, the public may expect risk to be reduced as much as possible, regardless of ROI. Our methodology simply supports the need to analyze the data; decision makers must then make the decisions.

COMPARING CASE STUDY RESULTS TO THOSE OF A CASE STUDY WITH IMPERFECT INFORMATION ASSUMPTIONS

Since we assumed perfect information for the results of our case study, we want now to show how these deterrence portfolios change for different attacker information imperfection biases. We show this analysis, using the same notional data as before but including each of the three attacker information imperfection biases: neutral, overconfident, and conservative. These biases are part of the scenario development, often referred to as "design-basis threat." This analysis is available from the authors, and has practical implications for decision makers. We summarize our limited analyses:

1. We get counterintuitive results for when a neutral attacker is assumed; we can actually show that when we have little to invest at the chemical facility (target A) it actually makes sense to make that information known to the attacker, because the average unconditional risk to the defender is less if the deterrence investment is known given this deterrence COA. In contrast, if we have a lot to invest at the chemical facility, we can show that it makes sense to withhold that information. These findings are due to the disparity in the notional maximum economic consequence values of the chemical facility and the ferry terminal, and may not hold if we change the notional target data. Thus, in circumstances similar to those of our notional example, a decision maker may prefer to make a counterintuitive decision and ensure perfect information if they deter by investing at the chemical facility, even if they do not invest much.
2. For when we assume an overconfident attacker, it makes sense for the defender to withhold information if their deterrence COA is to invest at a target with lower maximum economic value. This seems counterintuitive as in general if we have an overconfident attacker, making our deterrence information known to the attacker might reduce their confidence, but we show that this is not necessarily true. Thus, a decision maker in circumstances similar to those of our notional examples may prefer to invest at the chemical facility but ensure that the details of that investment are withheld from an overconfident attacker.
3. Finally, for when we assume a conservative attacker, it makes sense for the defender to communicate deterrence information if their deterrence COA is to invest at a target with a lower maximum economic value. This also seems counterintuitive as in general if we have a conservative attacker they might be emboldened by information about what the defender is doing to deter, but we show this is not necessarily true. Thus, a decision maker in circumstances similar to those of our notional examples may prefer to invest at the chemical facility, and should ensure the details of that investment are communicated to a conservative attacker.

DETERRENCE METHODOLOGY AND APPLICATION: SUMMARY

We have shown how to quantify deterrence effectiveness of different defender courses of action, based on the changes in attacker intent values after the defender deters, as compared to the attacker intent values before the defender deters. These intent values are functions of attacker pre-deterrence and post-deterrence expected utilities, and those utilities are functions of attacker capability to

attack, CI target vulnerability, and CI consequence if attacked. Together, capability, vulnerability, and consequence make up defender conditional risk, or the risk given an attack is already desired (intent=100%).

We have also shown how to quantify the average defender unconditional risk, or conditional risk multiplied by the attacker intent probability, both before and after measurable deterrence. Finally, we have shown how to calculate return on investment of the deterrence COA, and we have shown how to incorporate these data into a deterrence portfolio which CI officials and government agencies can use to report both quantifiable deterrence and resulting CI risk reduction/ROI. This approach allows a practitioner to implement DHS policy, which advocates both deterrence and risk reduction, in an integrated fashion. Our methodology is general enough to be applied to CI in multiple sectors, as our example demonstrated, and the general concepts of capability, vulnerability, and consequence can be applied to any type of intentional act.

Our example of how our methodology would be applied produced deterrence portfolios of four different deterrence COA. Overall, we found that investing in both targets in a two-target game yielded the best quantifiable deterrence effectiveness, the lowest unconditional risk to the defender, and the best ROI. Our brief analysis of a deterrence portfolio under imperfect information circumstances revealed some counterintuitive results with implications for decision makers.

DIRECTION FOR FUTURE WORK

Our methodology and findings are based on many assumptions. Future work on quantifying CI deterrence must consider these assumptions and vary them in order to discover new insights into this challenge. We offer the following specific suggestions.

CIKR SECTOR SPECIFICS

While we believe our proposed methodology is general enough to be applied to all DHS sectors, future work should examine ways to tailor this broad methodology to reflect nuances of CIKR in different CIKR sectors.

OWNER/OPERATOR ROLE

Future work may be necessary to modify this methodology for scenarios where owners/ operators implement security measures intended to deter, but bear the costs themselves.

INITIAL TARGET AND ACTOR DATA ASSUMPTIONS

We have assumed specific CI target characteristics, defender pre-deterrence investments in vulnerability and consequence reduction, attacker investment in capability enhancement, and other factors. Future work should modify this data before drawing conclusions from this methodology to quantify deterrence. Also, future work could change the complexity of various risk equation components as desired; for example, the consequence equations could be expanded to include factors such as loss of public trust.

UNCERTAINTY ANALYSIS

There is inherent uncertainty and subjectivity in probabilistic risk analysis. Since our methodology uses probabilistic capability, vulnerability, and consequence mitigation estimates, future work must address the uncertainty associated with these estimates and determine the implications for quantifiable deterrence effectiveness and resulting risk reduction. Tools such as MSRAM, which inform and are improved by our deterrence quantification methodology, have built-in subjectivity reduction measures that could be leveraged.

GAME THEORY APPROACH ASSUMPTIONS

We have used an unusual approach to a game theoretical analysis by discarding the notion of using either a pure or mixed strategy Nash Equilibrium. We also have taken the unusual approach of using attacker information imperfection biases as a proxy for

information when perfect information about defender investments is not available. Our approach is suited for our assumptions that neither pure nor mixed equilibrium strategies are appropriate for our present focus on quantifying the deterrence effects of defender investments. Nonetheless, future work should leverage more traditional game theoretical approaches to quantify deterrence. In support of such future work, we have already hypothesized how defender expected utility functions should be formulated for a more traditional game theoretical approach. The defender expected utility at equilibrium would be converted to loss, and then combined with attacker intent to attack (presumably with a pure NE that intent is 100%) to show the new defender unconditional risk.

Also, we assumed a two-target game; future work should examine how deterrence can be quantified in a game with more than two CI targets. This may require computer programs. Also, there are many repositories of CIKR data and risk information, such as MSRAM, ACAMS, and other databases to assist with this endeavor.

ADVERSARY TARGET PREFERENCES

Our example assumes the attacker has no preference for a certain target type; they simply want to maximize their overall expected utility from attacks. Future research can apply this methodology for issue-specific groups such as eco-terrorists who only want to target specific target types. This can support decision making on where to allocate defensive resources amongst similar target types in a defined geographic area.

OTHER RISK METHODOLOGIES

Our example has leveraged notional data from one risk analysis tool, the Coast Guard's MSRAM. Future work may use risk data from different tools and techniques to get the CIKR vulnerability and consequence data, and attacker capability data our approaches requires.

NETWORK ANALYSIS AND CIKR INTERDEPENDENCIES

Future work will examine how to quantify the deterrence effects of investments to secure networks of infrastructure from attack, including cyber networks. Network analysis uses different mathematical techniques to calculate network risk, but our methodology could still be applied. Future research could also consider the deterrent effects of investments in CIKR when those CIKR are co-located in or close proximity to other CIKR.

UTILITY FUNCTION COMPOSITION ASSUMPTIONS

We assumed attacker utilities and defender expected losses (risk) did not subtract their respective expenditures from the CI target values; the utility functions were only capability, vulnerability, and consequence modeled as functions of those expenditures. However, one could easily claim that attacker utility and defender risk must account for the respective lost expenditures, and subtract them from the expected benefit or add them onto the defender loss if an attack on a CI succeeds. Future work could create utility functions that subtract expenditures from target value. Also, we assumed the attacker's expected utility as a function of capability, vulnerability, and consequence, was the same as the defender's expected loss, thus creating a "symmetry" between attacker and defender functions. Introducing costs and budgets to the utility functions for each actor could create an asymmetrical game.

INFORMATION AVAILABILITY CIRCUMSTANCE ASSUMPTIONS

We have assumed that CI information was known to the attacker prior to defender deterrence efforts; future analysis could assume that the defender wants to quantify the deterrent effect of additional defensive investments even when their original CI investments and resulting security were not clear to the attacker. We have also assumed that attacker information is perfectly known

to the defender; in reality this is likely not true and should be examined.

In situations where perfect information is determined to be more advantageous, the CI operator must work actively with law enforcement and other government entities to aggressively communicate security and consequence mitigation deterrence measure information, in great detail to potential attackers. Obviously, this will be a challenging and imperfect endeavor, but the effort is still warranted if the methodology shows it is to our advantage to make information known to prospective attackers. The intelligence community could contribute to this effort as they glean information on what the adversary knows and believes about our deterrence efforts, but much work must be done by CI operators and law enforcement authorities, as well as public figures that communicate effectively and reach a broad audience.

However, if the effort required to develop and execute a public affairs deterrence campaign exceeds the expected loss to the government if deterrence fails, perhaps the campaign is not worthwhile, and the quantitative result supporting a perfect information advantage may be discarded.

In situations where imperfect information is determined to be more advantageous, the CI operator and law enforcement, as well as public figures, must collaborate to ensure security measures are not obvious to any observer. This could entail conduct "red cell" exercises where someone simulates a terrorist planning cell and visit a CI to see what security they can identify, and/or researches open source documentation on the security and economic value of CI. Also, the government may even have to classify security measures, such as federal grant funding.

UTILITY THEORY ASSUMPTIONS

We assumed subjective expected utility in our case study. In our initial work to expand this methodology leveraging prospect utility theory, we have used a linear scale for the PU utility and probability modifiers, but the work of Daniel Kahneman and Amos Tversky suggests a nonlinear scale.[45] Future work should leverage a nonlinear scale, and can exercise that scale by assuming different deterrence investment amounts in the deterrence game and determining which modifier applies. Kahneman and Tversky have done other work on decision analysis that could be applied to the utility functions. Also, Lisa Carlson and Raymond Dacey suggest that the PU probability weighting functions may determine the difference between behavior of attackers under SEU assumptions and under PU assumptions;[46] thus the extent to which deterrence might be more quantifiably effective under one assumption as compared to the other could be analyzed by explicitly observing the effects of different probability weights.

PROBABILITY-INVESTMENT RELATIONSHIP ASSUMPTIONS

Future work should also change the slope of the exponential probability-investment relationship curves that support capability, vulnerability, and economic consequence estimates. We assumed elimination fractions of 5% vulnerability and 5% capability to attack in our initial work to create deterrence portfolios leveraging exponential probability-investment relationship assumptions; these fractions could be modified to yield different curves and possibly yield different deterrence portfolio results. Also, we have assumed a linear relationship between target economic value retention and defender consequence mitigation investment in this paper's example. In reality, this relationship may be nonlinear. We only assumed that CI target vulnerability and attacker capability probability-investment relationships could be nonlinear on theoretical grounds; future work should analyze relationships between probability reduction and investments to develop rough cost curves. Finally, we have assumed attacker capability is reduced by defender offensive counterterrorism investment, but the attacker does not counteract this effect by increasing their investments. Future work should model the simultaneous effects of attacker and defender investment upon attacker capability.

ABOUT THE AUTHORS

Eric F. Taquechel *is a lieutenant commander in the United States Coast Guard and currently serves as chief of Contingency Planning and Force Readiness, Sector Boston. He has ten years of active duty commissioned Coast Guard service and previously served in USCGC GALLATIN, Marine Safety Office and Sector Houston-Galveston, and the Domestic Port Security Evaluation Division at USCG Headquarters. He most recently authored "Layered Defense: Modeling Terrorist Transfer Threat Networks and Optimizing Network Risk Reduction," in* IEEE Network Magazine. *LCDR Taquechel earned a master's degree in security studies from the Naval Postgraduate School and prior to that earned his undergraduate degree at the US Coast Guard Academy. LCDR Taquechel may be contacted at* eric.taquechel@uscg.mil.

Ted G. Lewis *is a professor of computer science and executive director of the Center for Homeland Defense and Security at the Naval Postgraduate School. He has forty years experience in academic, industrial, and advisory capacities, ranging from academic appointments at the University of Missouri-Rolla, University of Louisiana, and Oregon State University, to senior vice president of Eastman Kodak Company, to CEO and president of DaimlerChrysler Research and Technology, North America. Dr. Lewis has published over thirty books and 100 research papers. He is the author of* Critical Infrastructure Protection in Homeland Security: Defending a Networked Nation *(2006) and, most recently,* Network Science: Theory and Applications *(2009). He received his PhD in computer science from Washington State University. Dr. Lewis may be contacted at* tlewis@nps.edu.

ACKNOWLEDGEMENTS

The authors gratefully acknowledge the contributions of Dr. Richard Adler.

The views expressed herein are those of the authors and do not reflect the views of the United States Coast Guard or the Department of Homeland Security.

---

[1] U.S. Department of Homeland Security, *National Infrastructure Protection Plan* (2009), http://www.dhs.gov/xlibrary/assets/NIPP_Plan.pdf

[2] http://www.dhs.gov/xabout/laws/gc_1214597989952.shtm#1

[3] Andrew R. Morral and Brian A. Jackson, "Understanding the Role of Deterrence in Counterterrorism Security," RAND Occasional Paper (2009), http://www.rand.org/pubs/occasional_papers/2009/RAND_OP281.pdf

[4] See http://teamcore.usc.edu/projects/coastguard/

[5] Information about this initiative is available from the authors.

[6] http://birenheide.com/sra/2011AM/program/singlesession.php3?sessid=M4-H.

[7] In our example, we will show that deterrence measurement by itself may not be conclusive; it must be incorporated into a risk equation to have value and show changes in risk.

[8] DHS agencies currently are reporting risk reduction effectiveness of various CIKR security initiatives. For example, the Coast Guard reports port, waterways, and coastal security (PWCS) risk reduction efforts to Congress annually. Further, if DHS funds security investments for CIKR o/o, then they will necessarily be "joined at the hip" with those o/o to monitor implementation. Thus, DHS would be able to explain how the CIKR o/o's are deterring and thus reducing risk. For scenarios where DHS does not provide funding, there is no guarantee they could effectively report to Congress because they would not be coordinating with their CIKR stakeholders.

[9] Deterrence is the study of influencing an intelligent actor's decision making; this does not apply to unintentional threats where there is no human element or deliberative process.

[10] http://www.dhs.gov/xabout/laws/gc_1215444247124.shtm.

[11] Morral and Jackson, "Understanding the Role of Deterrence."

[12] Some DHS risk analysis tools, such as the US Coast Guard's Maritime Security Risk Analysis Model (MSRAM), have the ability to monetize human lives using a technique known as Value of Statistical Life (VSL), and draw equivalencies between different types of consequence using what is referred to as a Consequence Equivalency Matrix.

[13] The Coast Guard's MSRAM model uses a strategic threat value, without considering specific target attributes.

[14] Opportunistic attackers still face the challenge of overcoming security. If deterrence is influencing the decision making of an adversary, an opportunistic attacker could still be deterred by what they observe as they commence their attack, but that amounts to more of a tactical deterrence or last minute change in the attack planning cycle. With respect to those same security measures, more deliberate attackers would be deterred but earlier in their planning cycle, and have more time to adapt.

[15] Louis A. Cox, "Some Limitations of 'Risk=Threat x Vulnerability x Consequence' for Risk Analysis of Terrorist Attacks," *Risk Analysis 28* (2008): 1749-1761.

[16] Recoverability and resiliency measures may be too expensive for certain CIKR owners/operators to implement, but the risk equation still includes a placeholder for a consequence mitigation factor. In application, this mitigation factor may be 0, meaning expected consequence =100% of maximum potential consequence. Also, federal grants could help offset the costs of resilience measures.

[17] For a detailed explanation of MSRAM, see Brady C. Downs, "The Maritime Security Risk Analysis Model," *CG Proceedings 64* (2007): 36-38.

[18] DHS, *National Infrastructure Protection Plan.*

[19] Nikhil S. Dighe, Jun Zhuang, and Vicki M. Bier, "Secrecy in Defensive Allocations as a Strategy for Achieving more Cost Effective Deterrence," *International Journal of Performability Engineering* 5 (2009): 31- 43.

[20] Ibid.

[21]Morral and Jackson, "Understanding the Role of Deterrence."

[22] U.S. Department of Homeland Security, *DHS Risk Lexicon* (2010), http://www.dhs.gov/xlibrary/assets/dhs-risk-lexicon-2010.pdf.

[23] See, for example: Stanley Kaplan and John B. Garrick, "On the Quantitative Definition of Risk," Risk Analysis 1 (1981): 11-27; Tim Bedford and Roger Cooke, Probabilistic Risk Analysis: Foundations and Methods (Cambridge: Cambridge University Press, 2001); and Elisabeth Paté-Cornell and Seth Guikema, "Probabilistic Modeling of Terrorist Threats: A Systems Analysis Approach to Setting Priorities Among Countermeasures," *Military Operations Research* 7 (2002): 5-20.

[24] For examples of probabilistic risk tools being used in DHS, see Table 2-1 on pages 25-26 of the National Academy of Sciences Review of the Department of Homeland Security's Approach to Risk Analysis", 2010, available at http://www.nap.edu/openbook.php?record_id=12972&page=R1.

[25] Richard N. Lebow and Janet G. Stein, "Rational Deterrence Theory: I Think, Therefore I Deter," *World Politics* 41 (1989): 208–224.

[26] Elaine M. Bunn, "Can Deterrence Be Tailored?" *Strategic Forum*, No. 225 (Institute for National Strategic Studies, January 2007), http://www.hsdl.org/?view&did=481759.

[27] William M. McGill and Bilal M. Ayyub, "The Meaning of Vulnerability in the Context of Critical Infrastructure Protection," in *Critical Infrastructure Protection: Elements of Risk* (GMU School of Law Critical Infrastructure Protection Program, December 2007): 25-48, http://www.steelcityre.com/documents/RiskMonograph_1207.pdf.

[28] Patrick M. Cronin and Audrey K. Cronin, "Challenging Deterrence: Strategic Stability in the Twenty-first Century," *Special Joint Report of the International Institute for Strategic Studies and the University of Oxford Character of War Programme* (2007), http://ccw.modhist.ox.ac.uk/events/archives/mt06_deterrence/deterrence_report_mt2006.pdf.

[29] Richard N. Lebow, "The Cuban Missile Crisis: Reading the Lessons Correctly," *Political Science Quarterly* 98 (1983): 431–458.

[30] U.S. Department of Defense, *Deterrence Operations Joint Operating Concept*, version 2.0 (Washington, DC: Department of Defense, 2006), www.dtic.mil/futurejointwarfare/concepts/do_joc_v20.doc.

[31] John A. Major, "Advanced Techniques for Modeling Terrorism Risk," *Journal of Risk Finance* 4 (2002): 15-24.

[32] Bunn, "Can Deterrence by Tailored?".

[33] Erik Jenelius, Jonas Westin, and Åke J. Holmgren, "Critical Infrastructure Protection Under Imperfect Attacker Perception," *International Journal of Critical Infrastructure Protection* 3 (2010): 16-26.

[34] Martin J. Osborne and Ariel Rubenstein, *A Course in Game Theory* (Cambridge: MIT Press, 1994).

[35] Todd Sandler and Daniel G. Arce, "Terrorism & Game Theory," *Simulating and Gaming* 34 (2003): 319–337.

[36] Roger B. Myerson, *Force and Restraint in Strategic Deterrence: a Game Theorist's Perspective* (Strategic Studies Institute, US Army War College, 2007). http://www.strategicstudiesinstitute.army.mil/pubs/display.cfm?pubID=823.

[37] Morral and Jackson, "Understanding the Role of Deterrence."

[38] We use 95% elimination instead of 100% elimination to accommodate the theory that probability-investment relationships are exponential and thus probability of loss cannot be 100% eliminated, although in our case study examples our utility functions assume a linear relationship.

[39] Jenelius et. al., "Critical Infrastructure Protection Under Imperfect Attacker Perception."

[40] See http://www.fema.gov/pdf/government/grant/2011/fy11_psgp_kit.pdf for more information about Port Security Grants.

[41] The US Coast Guard's MSRAM tool includes maritime CIKR in all eighteen DHS CIKR sectors.

[42] Computing power would likely also be necessary for quantifying deterrence effects of investments, when multiple attack modes for the targets under consideration are modeled. Our example only assumes one attack mode (boat bomb).

[43] Port Security Grants could be used to reduce consequences as well, to add deterrent value. Overseas military operations and counterterrorism actions are examples of actions that mitigate attacker capabilities.

[44] Across all attacker options means the attacker uses the same attack mode (boat bomb), but could attack different combinations of CIKR. Since we don't use the traditional Nash Equilibrium to find the optimal attacker choice after we invest to deter, we need to aggregate resulting risk across all attacker choices.

[45] Daniel Kahneman and Amos Tversky, "The Framing of Decisions and the Psychology of Choice," *Science* 211 (1981): 453-458.

[46] Lisa J. Carlson and Ramond Dacey, "Sequential Analysis of Deterrence Games with a Declining Status Quo," *Conflict Management and Peace Science* 23 (2006): 181-198.

Copyright © 2012 by the author(s). *Homeland Security Affairs* is an academic journal available free of charge to individuals and institutions. Because the purpose of this publication is the widest possible dissemination of knowledge, copies of this journal and the articles contained herein may be printed or downloaded and redistributed for personal, research or educational purposes free of charge and without permission. Any commercial use of *Homeland Security Affairs* or the articles published herein is expressly prohibited without the written consent of the copyright holder. The copyright of all articles published in *Homeland Security Affairs* rests with the author(s) of the article. *Homeland Security Affairs* is the online journal of the Naval Postgraduate School Center for Homeland Defense and Security (CHDS).

http://www.hsaj.org

INTECH
open science | open minds

OPEN ACCESS
ARTICLE

International Journal of Advanced Robotic Systems

Game Theory Models for Multi-Robot Patrolling of Infrastructures

Regular Paper

Erik Hernández[1,*], Jaime del Cerro[1] and Antonio Barrientos[1]

1 Center for Robotis and Automation, UPM-CSIC, Madrid. Spain
* Corresponding author E-mail: ehernandez@industriales.upm.es

Received 14 Jun 2012; Accepted 17 Dec 2012

DOI: 10.5772/55535

© 2013 Hernández et al.; licensee InTech. This is an open access article distributed under the terms of the Creative Commons Attribution License (http://creativecommons.org/licenses/by/3.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Abstract This work is focused on the problem of performing multi-robot patrolling for infrastructure security applications in order to protect a known environment at critical facilities. Thus, given a set of robots and a set of points of interest, the patrolling task consists of constantly visiting these points at irregular time intervals for security purposes. Current existing solutions for these types of applications are predictable and inflexible. Moreover, most of the previous work has tackled the patrolling problem with centralized and deterministic solutions and only few efforts have been made to integrate dynamic methods. Therefore, one of the main contributions of this work is the development of new dynamic and decentralized collaborative approaches in order to solve the aforementioned problem by implementing learning models from *Game Theory*. The model selected in this work that includes belief-based and reinforcement models as special cases is called *Experience-Weighted Attraction*. The problem has been defined using concepts of *Graph Theory* to represent the environment in order to work with such *Game Theory* techniques. Finally, the proposed methods have been evaluated experimentally by using a patrolling simulator. The results obtained have been compared with previous available approaches.

Keywords Multi-Robot Patrolling, Game Theory, Experience-Weighted Attraction, Security Application

1. Introduction

Domains, where distributed surveillance, inspection or control are required, are candidates for being secured by performing patrolling tasks, usually by walking throughout at regular intervals (Abate, 1996), (Almeida et al., 2004). Currently, security system solutions are mostly predictable and inflexible. Additionally, since they are controlled by human operators, their performance can be affected by limitations such as boredom, distraction, or fatigue. Furthermore, in some environments, people must deal with hazardous conditions. As a consequence, it is important to improve the security elements used in these types of systems, which assist human beings in dangerous scenarios such as mine clearing or search and rescue operations. They are then able to perform other type of high level tasks, i.e., monitoring the system from a safe location (Oates et al., 2009). Recently, new research efforts have arisen trying to solve some challenges related to security tasks automation by using mobile robots (Everett, 2003). Thus, mobile robots aim to perform some useful task that a human either cannot, or

would prefer not to do. Moreover, the robot should hopefully do it better, cheaper, safer, and more reliably.
Security systems that utilize mobile robots in these types of applications have a great deal of advantages, i.e., they do not experience human limitations. However, some tasks are too complex that a single robot cannot achieve good results, especially in the presence of uncertainties, incomplete information, distributed control, and so forth. To overcome these challenges, *Multi-Robot Systems* can be used. They are characterized as a set of homogeneous or heterogeneous robots operating in the same environment using cooperative behaviors (Farinelli et al., 2004).

In this paper, new collaborative multi-robot approaches for infrastructure security applications at critical facilities are explored. The work is focused on *area patrol*, i.e., the activity of going throughout an area. Thus, given a set of robots and a set of points of interest, the patrolling problem consists of constantly visiting these points at irregular time intervals for security purposes. This problem has been formulated using concepts of *Graph Theory* to represent an environment where nodes stand for specific location of interest and edges for possible paths. By using this representation, each path has a cost that represents the time required to go from one node to another. The main advantage of this representation is its application. It can be easily used in other domains, i.e., computer networks, distributed coverage, and so forth. Additionally, there is a wide variety of problems that may be reformulated as particular patrolling task such as cleaning or surveillance. Since the patrolling problem seeks to maximize the number of visits to each node in a given environment, a good patrolling strategy must reduce the time lag between two visits to the same location (Chevaleyre, 2004).

The main contributions of this work are summarized as follows: An analysis of the behavior of game theory models in the multi-robot patrolling problem context is presented. A dynamic and distributed solution has been developed in order to solve the aforementioned problem. A novel decision-making rule has been defined. This rule attempts to allow robot dispersion, i.e., at each point of interest, each robot chooses a different available set of actions. A demonstration of how multiple robot interaction arises with the definition of multiple games at each point of interest has been defied. Finally, a detailed study of the behavior of the implemented model parameters has been described.

The rest of this paper is organized as follows. Section 2 briefly describes related work. Section 3 gives definitions of game theory and introduces the problem. Section 4 shows the implemented models in order to solve the patrolling problem. Section 5 presents the evaluation and experimental results. Finally, section 6 summarizes the obtained results.

2. Related work

The multi-robot patrolling problem has received much attention in recent years, specially in works that develop algorithms to coordinate decision-making among robots, (Portugal and Rocha, 2011). These works have implemented different principles such as reinforcement learning (Santana et al., 2004); negotiations methods (Hwang, 2009); swarm optimization (Glad and Buffet, 2009); cycle and partitioning strategies (Chevaleyre, 2004); and adaptive solutions (Sempé and Drogoul, 2003). A description of all of them can be found on a recent survey by (Portugal and Rocha, 2011). Beyond this survey, the multi-robot patrolling problem was tackled in (Ahmadi and Stone, 2006). In that work, the problem was called *Continuous Area Sweeping*, which is solved with a partitioning area method. Moreover, in (Aguirre et al., 2011), the multi-robot patrolling is applied to patrol national borders. In that work, elements of game theory as well as Monte Carlo simulation are used to solve the problem via genetic algorithms. Another work that utilizes game theory principles is described in (An et al., 2012). In that work, solutions to solve competitive or zero-sum games for the protection of critical infrastructure via *Stackelberg Games* are presented.

Among all these works, three of them are directly related with this work. The pioneer work in the multi-robot patrolling problem was carried out by (Machado et al., 2003). In that work, authors defined an evaluation criterion based on idleness. Idleness is the time that a place remains unvisited. Thus, *Total Idleness* is defined as the average of the idleness of all places of a given environment. Since this criterion is widely used in literature, it was used to measure the performance of the methods proposed in this work. Moreover, the problem of generating a patrol path inside a target area was tackled in (Elmaliach et al., 2007). The algorithm applied to generate this patrol path is called *Cycle* and it guarantees that each point is covered with the same optimal frequency. The solution presented in that work uses Spanning Tree Coverage method to find a minimal Hamilton path of minimal costs. Once a path is obtained, robots are uniformly distributed along this path and follow the same patrol route over and over. Thus, uniform frequency of the multi-robot patrolling task is achieved as long as one robot continues working properly. Moreover, authors present criteria based on frequency optimization in order to evaluate multi-robot patrolling algorithms. Finally, in (Portugal and Rocha, 2010) is presented an algorithm called *MPS*. Such an algorithm divides the environment into regions with the same dimension by using a balanced graph partitioning approach. Each of these regions is assigned to a robot that follows a local patrolling route. The procedure to

obtain this patrolling route mainly seeks Euler and Hamilton circuits and paths. However, if such circuits and paths do not exist, the procedure seeks longest paths and Non-Hamiltonian cycles. Non-Hamiltonian cycles are selected only when they have at least half of the vertices of a graph; if not, the patrolling route remains the longest path. Since the longest path and the Non-Hamiltonian cycle do not contain all vertices of the graph, the procedure includes such vertices to complete the patrolling route. Then, ultimately inverse path procedure is used to return to the starting vertex of the route when is required.

Previous literature has demonstrated the effectiveness of methods that implement solutions based on cycles and paths (Chevaleyre, 2004). The good performance of these approaches could be explained by their centralized and explicit coordinator scheme, (Almeida et al., 2004). However, a centralized solution has several disadvantages such as lack of scalability in the number of places to protect and susceptibility to single-point failure, due to its unique, and hence vulnerable, control point. In addition, these approaches are deterministic, and therefore not suitable for security purposes due to their predictability.

The present work differs from others on the manner in which the patrolling problem was solved by implementing learning models from *Game Theory*. The theory of learning in games defines equilibrium as the result of dynamic adjustment processes in which players interact for optimality over time in repeated normal-form games. Thus, they compute their *myopic* best response based on the accumulated experience achieved by tracking previous plays history of other players. The learning model selected in this work to patrol throughout an environment was proposed by Camerer and is called *Experience-Weighted Attraction* (Camerer, 1999). Implementing such adaptive models allows developing dynamic and distributed solutions similar to (Santana et al., 2004) in contrast to several literature works.

3. Concepts from game theory

A brief overview of concepts as well as some definitions of game theory (Fudenberg, 1998) are given in order to clarify the description in the following sections. In this work, an abstract representation of the environment as an undirected weighted graph G has been adopted. This graph is an ordered pair consisting of a set $E(G)$ of edges and a set $N(G)$ of nodes. Each node is a special point of interest that needs to be observed in search of intruders, but it is assumed that such observation is instantaneous. Each edge represents a path by using a number corresponding to the cost proportional to its length.

Thus, given such graph and a set of robots, the patrolling task consists of visiting at each time step as many nodes as possible in order to minimize the time lag between two visits at the same node. Therefore, each node not only is an environment point of interest to be inspected, but also a point where interaction among agents arises, i.e., each robot in graph node $n \in N(G)$ must select, based on other robots selections, an appropriate action in order to choose the next node to visit. Taking into account this interaction, normal-form games at each graph node $n \in N(G)$ have been defined.

Definition 1 (Normal-Form Game) Formally, a finite n-robot normal-form game Γ is made of:

- A finite set M of *robots* $i = 1,\ldots,n$.
- A finite set $A = A_1 \times \cdots \times A_n$, where $A_i = \{a_i^1,\ldots,a_i^k\}$ is a finite set of actions for *robot* $i = 1,\ldots,n$. Each vector $a = \{\{a_1^{j_1},\ldots,a_n^{j_n}\} \in A \mid j_{1,\ldots,n} \in \{1,\ldots,k\}\}$ is called *action profile* for the game Γ. Each action is related to an edge $e \in E(G)$.
- A finite set $S = S_1 \times \cdots \times S_n$, where $S_i = \{s_i^1,\ldots,s_i^k\}$ is a finite set of strategies for *robot* $i = 1,\ldots,n$. Each vector $s = \{\{s_1^{j_1},\ldots,s_n^{j_n}\} \in S \mid j_{1,\ldots,n} \in \{1,\ldots,k\}\}$ is called *strategy profile* for the game Γ. A strategy is the criterion taken into account to determine the action to be selected.
- A payoff function $\pi_i(s) \mapsto \Re$ for *robot* $i = 1,\ldots,n$, where S is the set of strategy profiles.
- Therefore, $\pi_i(s)$ is the payoff of robot i when strategy profile s is chosen.

In this implementation, each node $n \in N(G)$ has a fixed number of normal-form games which depends on the edges connected to it. At time step zero, the robots are randomly positioned on the set $N(G)$.

Thus, at every time step, the robot $i \in M$ reaches a node $n \in N(G)$ and plays its corresponding normal-form game Γ. As a consequence, the robot chooses its individual strategy $s_i^{j_i} \in S_i$ considering the strategies selected by all other robots. The action related to the strategy chosen leads the robot to the next node. Finally, the interaction among robots arises when each robot sends a message indicating the strategy selected. A robot can select an action with probability one or by randomizing over the set of available actions according to some probability distribution. Such strategies are called *pure* and *mixed*, respectively.

Definition 3 (Pure Strategy) Given a set of available actions $A_i = \{a_i^1, \ldots, a_i^k\}$ for robot $i = 1, \ldots, n$, a pure strategy $s_i^{l_i} = \{a_1^{j_1} | j_i \in \{1, \ldots, k\}\}, l_i \in \{1, \ldots, k\}$. Thus, given a game Γ, a *pure strategy profile* $s = \{\{s_1^{j_1}, \ldots, s_n^{j_n}\} | j_{1,\ldots,n} \in \{1, \ldots, k\}\}$ for the game is the joint strategies selection where $s_n^{j_n}$ is the choice of strategy j_n by robot $n \in M$

Definition 4 (Mixed Strategy) A mixed strategy σ_i for robot $i = 1, \ldots, n$ is a probability distribution over the set of available actions $A_i = \{a_i^1, \ldots, a_i^k\}$ for such robot; i.e., a *mixed strategy* has the form $\sigma_i = (p_i^1 a_i^1 + \cdots + p_i^k a_i^k)$ where $p_i^{j_i} \geq \{0 | j_i = 1, \ldots, k\}$ and $\Sigma_{j_i=1}^k p_i^{j_i} = 1$. Thus, robot $i \in M$ selects $a_i^{j_i}$ with probability $p_i^{j_i}$. If $p_i^{j_i} = 0, \forall j = 1, \ldots, k$ except for one, say $p_i^{l_i} = 1, l_i \in \{1, \ldots, k\}$, then σ is a *pure strategy*. Thus, given a game Γ, a *mixed strategy profile* $\sigma = (\sigma_1, \ldots, \sigma_n)$ for the game is the joint selection of strategies where σ_i is the strategy selected by robot $i \in M$.

A manner to represent a game Γ is through the use of a *n × n matrix payoff game* in which each cell of this matrix indicates a strategy profile and contains the outcome or *mixed strategy payoffs* of the game when such profile is selected. Such outcome take the form $C(x, y) = (\pi_1, \ldots, \pi_n), x, y = 1, \ldots, n$ where the entry π_i is the payoff of robot $i \in M$

Definition 5 (Expected payoff of a mixed strategy) Given a normal-form game Γ, the expected payoff for robot $i \in M$ playing the strategy profile for the game $s = (s_1^{j_1}, \ldots, s_n^{j_n}), j_{1,\ldots,n} \in \{1, \ldots, k\}$ is defined as

$$\pi_i(s) = \left\{ \sum_{a \in A} \pi_i(a) \cdot \prod_{i=1}^{n} \Pr(a_i^{j_i}) | j_i \in \{1, \ldots, k\} \right\} \quad (1)$$

where $\Pr(a_i^{j_i})$ is the probability that action j_i will be played by robot $i \in M$.

Thus, the robots that interact in these types of games choose an action that maximize its expected payoff considering the actions selected by all other robots. This is called *best response* and it leads to the central solution concept of game theory, the *Nash equilibrium*. From now on, the robots other than robot $i \in M$ are specified by $-i = \{(1, \ldots, i-1, i+1, \ldots, n) \in M\}$. Moreover, the strategy profile without the strategy of robot $i \in M$ is defined by

$$s_{-1} = \left\{ (s_1^{j_1}, \ldots, s_{i-1}^{j_{i-1}}, s_{i+1}^{j_{i+1}}, \ldots, s_n^{j_n}) \in S_i | j_{1,\ldots,i-1,i+1,\ldots,n} \in \{1, \ldots, k\} \right\}.$$

Definition 6 (Best Response) The best response of robot $i \in M$ to the strategy profile s_{-1} is a strategy $s_i^* \in S_i$ such that $\pi_i(s_i^*, s_{-i}) \geq \pi_i(s_i^{j_i}, s_{-i}), j_i \in \{1, \ldots, k\}$ for all strategies $s_i^{j_i} \in S_i$.

Definition 7 (Nash equilibrium) A strategy profile $s = \{(s_1^{j_1}, \ldots, s_n^{j_n}) | j_{1,\ldots,n} \in \{1, \ldots, k\}\}$ for the game Γ is a Nash Equilibrium if, for all robots $i \in M, s_i^{j_i}$ is a best response to s_{-i}.

In the games played in this implementation, the robots do not have conflicting interests and their sole challenge is to coordinate on actions that are maximally beneficial to all. These types of games are called *team games*.

Finally, in order to visit as many nodes of a graph as possible, a set of robots must disperse throughout the environment. Based on this requirement, the payoffs are defined as follow: let $\tau_i(s_{-i}^j)$ be the times that robots other than robot $i \in M$ select the strategy $j \in \{1, \ldots, k\}$, where $s_{-i}^j = \{(s_1^j, \ldots, s_{i-1}^j, s_{i+1}^j, \ldots, s_n^j)\}$. Thus, the payoff for robot $i \in M$ playing such strategy is defined as $\pi_i(s_i^j, s_{-i}^j) = |M| - \tau_i(s_{-i}^j)$. Therefore, the payoff is low when several robots choose the same strategy. Henceforth, at each time step that a robot $i \in M$ reaches a node $n \in N(G)$, plays a normal-form game Γ against $-i \in M$, chooses an available action of its set of actions A_i and goes to the next node $n \in N(G)$.

4. Experience-weighted attraction learning model

In the *Experience-Weighted Attraction* or *EWA* learning model each strategy $s_i^{j_i} \in S_i$ has a numerical value called *attraction*, which specifies the probability of choosing that strategy. Each attraction has an initial value, which is updated each period through the use of two rules that update two variables. The first variable $\Lambda_i^{j_i}(t)$, correspond to the level of attraction of the robot $i \in M$ by the strategy $s_i^{j_i} \in S_i$ after period t. The second variable $\psi(t)$ represents the amount of experience and indicates the number of observation-equivalents of past experience. These variables begin with some prior values, $\psi(0)$ and $\Lambda_i^{j_i}(0)$.

Thus, the first rule updates the level of attraction according to

$$\Lambda_i^{j_i}(t) = \frac{\phi \cdot \psi(t-1) \cdot \Lambda_i^{j_i}(t-1) + \left[\delta + (1-\delta) \cdot \xi(S_i, s_i^{j_i}(t))\right] \cdot \pi(s_i^{j_i}(t), s_{-i})}{\psi(t)} \quad (2)$$

The decay rate ϕ depreciates previous attraction $\Lambda_i^{j_i}(t-1)$ and represents a combination of forgetting and level of knowledge about the adaptation of other robots. When ϕ

is lower, old attractions are decayed more quickly, whereas the most recent attractions are more important. Moreover, the parameter δ weights hypothetical payoffs that unchosen strategies would have earned, whereas $(1-\delta)$ weights payoffs received from the chosen strategy $s_i^{j_i}(t)$. Finally, the indicator function $\xi\left(S_i, s_i^{j_i}(t)\right)$ is equal to 1 when the strategy selected $s_i^{j_i}(t)$ at period t is equal to some strategy of the set of strategies S_i. This indicator function is utilized in all the models of this section.

The second rule updates the amount of experience according to

$$\psi(t) = (1-\nu)\cdot\phi\cdot\psi(t-1)+1, t \geq 1 \tag{3}$$

where $(1-\nu)\cdot\phi$ is the rate of decay for experience, which measures the impact of previous experience. Moreover, the parameter ν specifies how quickly robot $i \in M$ lock in to a strategy. When $\nu = 0$, attractions are weighted averages of past attractions and payoffs. Past attractions are weighted by $\frac{\phi\cdot\psi(t-1)+1}{\phi\cdot\psi(t-1)+1}$, whereas payoffs are weighted by $\frac{1}{\phi\cdot\psi(t-1)+1}$. When $\nu = 1$ the attractions cumulate because $\psi(t)$ becomes one.

The initial attraction level of each strategy $s_i^{j_i} \in S_i$ is denoted by $\Lambda_i^{j_i}(0) = \frac{1}{|S_i|}, \forall j_i = 1,\ldots,k$, while the initial value for the number of *observation-equivalents* $\psi(0) = 1$.

Thus, an EWA robot $i \in M$ using these rules shapes a set of attractions which specify preference for a specific strategy $s_i^{j_i} \in S_i$. Such preference is given as a choice probability $P_i^{j_i}$ in time step $t+1$ through a logistic stochastic response function defined by

$$P_i^{j_i}(t+1) = \frac{e^{\lambda\cdot\Lambda_i^{j_i}(t)}}{\sum_{j_i=1}^{k} e^{\lambda\cdot\Lambda_i^{j_i}(t)}} \tag{4}$$

where λ is the response sensitivity. With $\lambda = 0$ the choice is stochastic while $\lambda = \infty$ is best response.

Beyond these rules, specific values of $\psi(0), \delta$ and ρ reduce this general model to special cases such as *reinforcement* and *belief-based* models.

4.1 Reinforcement Model

In the *reinforcement* model of EWA, every time step that a robot $i \in M$ reaches a graph node $n \in N(G)$, it performs three steps. In the first step, the robot selects one of the strategies available at such node. This selection is based on a logistic stochastic response function defined by

$$P_i^{j_i}(t+1) = \frac{e^{\lambda\cdot R_i^{j_i}(t)}}{\sum_{j_i=1}^{k} e^{\lambda\cdot R_i^{j_i}(t)}} \tag{5}$$

where λ is the response sensitivity. With $\lambda = 0$ the choice is stochastic, whereas $\lambda = \infty$ is best response. The response function $P_i^{j_i}$ specifies the selection probability of the strategy $s_i^{j_i}$

Each strategy $s_i^{j_i}$ is related to a reinforcement value $R_i^{j_i}$. When this reinforcement value is updated, its related strategy is reinforced. Thus, in the second step, once a strategy is selected, only this strategy is reinforced by previous received payoffs according to

$$R_i^{j_i}(t) = \phi\cdot R_i^{j_i}(t-1) + \xi\left(S_i, s_i^{j_i}(t)\right)\cdot\pi_i\left(s_i^{j_i}(t), s_{-i}\right) \tag{6}$$

As can be seen, this rule is the result of setting in the EWA model $\delta = 0, \nu = 1$, and $\psi(0) = 1$, therefore $\psi(t) = 1$. This is the simplest form of cumulative reinforcement. When $\delta = 0, \nu = 0$ and $\psi(0) = \frac{1}{(1-\phi)}$, *EWA* model becomes a form of averaged reinforcement governed by

$$R_i^{j_i}(t) = \phi\cdot R_i^{j_i}(t-1) + (1-\phi)\cdot\xi\left(S_i, s_i^{j_i}(t)\right)\cdot\pi_i\left(s_i^{j_i}(t), s_{-i}\right) \tag{7}$$

where reinforcements are averages of previous attractions and incremental reinforcement. The initial reinforcement value of the strategies available at each node is defined by $R_i^{j_i}(0) = \frac{1}{|S_i|}, \forall j_i = 1,\ldots,k$.

Finally, in the last step, the robot $i \in M$ communicates the strategy selected to the other robots, so that they update the reinforcement value of the strategy selected by the robot $i \in M$ in the node $n \in N(G)$. Thus, similar to the behavior of attractions in EWA model, in the *reinforcement* case, each robot shapes the reinforcement of each strategy by utilizing the aforementioned rules. The Algorithm 1 describe the three steps accomplished in this model.

Data: M, N
forall $n \in N(G)$ **do**
  forall $j \in 1,\ldots,k$ **do**
    $R^j(0) \leftarrow \frac{1}{|S|}$
  end
end
while *true* **do**
  (1) Select a strategy based on P^j;
  (2) Reinforce the strategy selected through R^j;
  (3) Communicate the strategy selected to the other robots $- i \in M$;
end

Algorithm 1. Algorithm of the reinforcement model

4.2 Belief-Based Model

Belief-based models start with the premise that each robot $i \in M$ identifies that it is playing a game Γ with other robots, and it forms beliefs about what these robots will play in the future based on its past observation. Then, it attempts to define dynamic processes that lead to a *Nash Equilibrium* by choosing a best response strategy that maximizes its expected payoff to its beliefs.

There are different iterative learning rules to form beliefs. One widely used model of learning is the process of *Weighted Fictitious Play* and its variants, such as *Cournot Best-Response Dynamics*, which looks back only one play, as opposed to *Fictitious Play* which looks back the t most recent plays, (Brown, 1951). At each time step in the model of *Weighted Fictitious Play*, each robot $i \in M$ chooses its strategies to maximize its expected payoff given its prediction about the distribution over strategies of other robots at that time step. Therefore, *Weighted Fictitious Play* is an instance of model-based learning in which a robot maintains beliefs $B_i^t(s_{-i})$ about the strategies of other robots $s_{-i} \in S_{-i}$. In the prediction of this learning rule, the initial prior belief that robot $i \in M$ assigns to strategies s_{-i} of robots $-i \in M$ is governed by

$$B_i^0(s_{-i}) = \frac{Y_i^0(s_{-i})}{\sum_{\tilde{s}_{-i} \in S_{-i}} \left[Y_i^0(\tilde{s}_{-i}) \right]} \tag{8}$$

where $Y_i^0(s_{-i}) : S_{-i} \to \Re_+$ is an exogenous initial weight function, which assigns a real value to each strategy of the robots other than robot $i \in M$. This value is assigned according to $Y_i^0(s_{-i}) : \frac{|S_{-i}| - (j_i - 1)}{|S_{-i}|}, \forall j_i = 1, \ldots, k,$ where $|S_{-i}|$ represents the cardinality of the set S_{-i}. Later, every time step that the strategy s_{-i} is played, its weight is updated according to

$$Y_i^t(s_{-i}) = \phi \cdot Y_i^{t-1}(s_{-i}) + \begin{cases} 1 & \text{if } s_{-i}^{t-1} = s_{-i} \\ 0 & \text{if } s_{-i}^{t-1} \neq s_{-i} \end{cases} \tag{9}$$

The initial weight assigned is different for each strategy. This assignation permits that the updates performed by 9 do not lead to weights with the same value, which allows to avoid selection problems.

The belief that robot $i \in M$ assigns to the robots $-i \in M$ playing s_{-i} at time step t is given by

$$B_i^t(s_{-i}) = \frac{Y_i^t(s_{-i})}{\sum_{\tilde{s}_{-i} \in S_{-i}} \left[Y_i^t(\tilde{s}_{-i}) \right]} \tag{10}$$

The updating rule formulated by 10 can be defined in terms of previous-period beliefs by

$$B_i^t(s_{-i}) = \frac{\phi \cdot \sum_{\tilde{s}_{-i} \in S_{-i}} \left[Y_i^t(s_{-i}) \right] \cdot B_i^{t-1}(s_{-i}) + \begin{cases} 1 & \text{if } s_{-i}^{t-1} = s_{-i} \\ 0 & \text{if } s_{-i}^{t-1} \neq s_{-i} \end{cases}}{\phi \cdot \sum_{\tilde{s}_{-i} \in S_{-i}} \left[Y_i^t(\tilde{s}_{-i}) + 1 \right]} \tag{11}$$

In 11, $B_i^t(s_{-i})$ is expressed in terms of previous-period beliefs, similar to EWA model with previous-period attractions and reinforcement model with previous-period reinforcements.

Following this updating rule, most recent and previous-period beliefs are updated ϕ times. When $\phi = 0$ *Weighted Fictitious Play* becomes *Cournot Best-Response Dynamics*, and when $\phi = 1$ it becomes *Fictitious Play*. Once beliefs are updated, expected payoff of robot $i \in M$ in period t, $E_i^t(s_i^{j_i})$ is defined according to

$$E_i^t(s_i^{j_i}) = \sum_{\tilde{s}_{-i} \in S_{-i}} \pi(s_i^{j_i}(t), \tilde{s}_{-i}) \cdot B_i^t(s_{-i}) \tag{12}$$

As in the case of beliefs, expected payoffs can be expressed as a function of previous-period expected payoffs which yields

$$E_i^t(s_i^{j_i}) = \frac{\phi \cdot \sum_{\tilde{s}_{-i} \in S_{-i}} \left[Y_i^t(\tilde{s}_{-i}) \right] \cdot E_i^{t-1}(s_i^{j_i}) + \pi(s_i^{j_i}(t), s_{-i})}{\phi \cdot \sum_{\tilde{s}_{-i} \in S_{-i}} \left[Y_i^{t-1}(\tilde{s}_{-i}) \right] + 1} \tag{13}$$

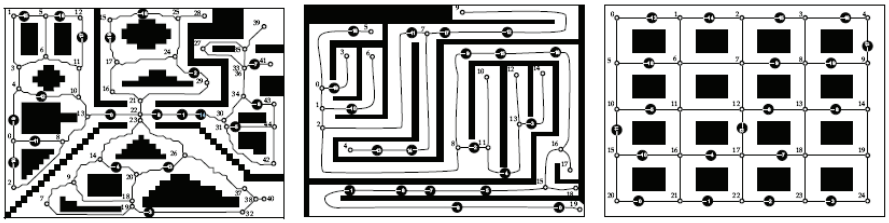
Finally, the best response of the robot $i \in M$ in *Weighted Fictitious Play* is given by

$$BR_i^t = \arg\max_{j_i} E_i^t(s_i^{j_i}) \tag{14}$$

5. Experiments and result

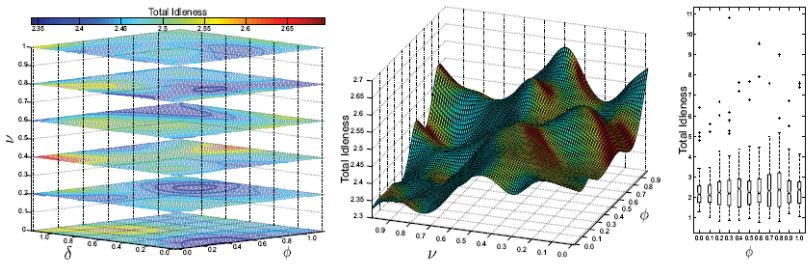
In order to evaluate and compare this implementation with other methods, a patrolling simulator developed from pioneers works (Machado et al., 2003) has been used.

Thus, the first experiments aim at analyzing the behavior of these models with different values of their parameters, namely, for *EWA* model ϕ, δ and ν, for reinforcement model ϕ and ν, and for belief-based model ϕ. In order to do so, the map shown in figure 1(a) was used. Where unfilled small circles stand for nodes or points of interest, lines stand for edges of a graph or paths that robots use to move throughout the map. Filled big circles stand for robots patrolling such map. In this set of experiments, a group of 20 robots started at node number 22 and patrol until each node had been visited 256 times.



(a) Strongly Connected Map (b) Maze Map (c) Grid Map

Figure 1. Maps used in order to evaluate performance of each algorithm



(a) *EWA* (b) Reinforcement (c) Belief-based

Figure 2. Performance of the models evaluated utilizing the map of figure 1(a).

Figure 2 shows the performance of the models evaluated. At each plot, color intensity, coordinates, and box plots represent the total idleness. Thus, figure 2(a) shows the behavior of *EWA* model using six slice planes at the axis ν through a volumetric data created with values of $0 \le \nu, \phi, \delta \le 1.0$. As it can be seen, the best behavior was obtained with the slice at $\nu = 0.2$. In this slice, the volumetric data shows that the idleness is, in general, approximately 2.5. The region around the points $\delta = 0.7$ and $\phi = 0.9$ presents the lowest idleness, whereas the idleness value hardly ever reaches values of 2.55. Figure 2(b) shows a surface expressing the behavior of the reinforcement model for values of $0 \le \nu, \phi, \delta \le 1.0$. It shows that the bigger the value of ν and the lower the value of ϕ, the better the performance of the model. Indeed, the surface has a minimum at coordinate $(\nu = 0.9, \phi = 0.3)$. In general, $\nu = 0.9$ achieves the best behavior, which indicates that the form of cumulative reinforcement performs better than the average reinforcement. In the form of cumulative reinforcement, each strategy has a level of reinforcement, which is incremented cumulatively by received payoffs. Finally, figure 2(c) shows the performance of the belief-based model for values of $0.0 \le \phi \le 1.0$ by means of using box plots. The size of each box represents the spread of the data, whereas the symbols "+", called outliers, are values distant from the rest of the data. For security purpose, it is not suitable to have several outliers because they represent nodes with high idleness, which indicates points of weakness. In general, the nature of the belief-based model is almost the same in all cases. When $\phi = 0.1$ the median, 2.29, of the data is the smallest. Even though, the smallest median does not mean better performance, the size of the box when $\phi = 0.1$ is lower than the other options. This size indicates that the total idleness of all nodes is similar. Moreover, there are only two outliers, at nodes 28 and 40 with total idleness of 5.22 and 5.61, respectively. Although, there are options with two or less outliers, their values are bigger. Based on this information, values of $\nu = 0.2, \phi = 0.9$ and $\delta = 0.7$ for *EWA* model; $\nu = 0.9$ and $\phi = 0.3$ for reinforcement model, and $\phi = 0.1$ for belief-model model were selected.

Once all parameters of the models were determined, the next experiments aimed at evaluating the models in comparison to other methods available in the Patrolling Simulator, namely, the *Cycle* algorithm and the *MPS* algorithm, which were described in section 2. For these experiments, all the maps of figure 1 were used. For each map, sets M of 6, 8, 10, 15, 20, 25, 30 robots were evaluated. This combination generates 21 cases of study to experiment. Each experiment consists of visiting all nodes 256 times. At time $t = 0$, each robot was placed randomly at one node of the map.

Table 1 shows the results of these experiments for *Strongly Connected*, *Maze* and *Grid* maps. In these results, the lower the size of μ, the better the performance of the algorithm and the lower the value of σ, the lower the variance. As it can be observed, if the value of M increases, the performance of all of the algorithms is better. Although this behavior seems obvious, if the coordination among robots is not appropriate, bigger values of M will not lead to better performance of the algorithms. Therefore, this nature is due to the suitable coordination.

Map	M	Cycle		MPS		Belief-Based		Reinforcement		EWA	
		Total Idleness									
		μ	σ	μ	σ	μ	σ	μ	σ	μ	σ
Strongly Connected	30	1.711	0.829			1.663	0.649	**1.647**	0.614	1.674	0.673
	25	2.006	0.974			**1.904**	0.655	2.086	1.109	1.979	0.876
	20	2.471	1.195			**2.445**	1.151	2.460	1.173	2.453	1.101
	15	2.981	3.805			3.066	2.672	**2.372**	2.923	3.598	4.981
	10	5.119	6.534			3.941	5.090	**3.148**	2.640	5.587	5.165
	8	**4.225**	3.662	5.889	5.287	4.912	4.095	4.443	4.714	5.130	4.714
	6	8.139	10.81	6.783	6.689	8.474	9.503	8.078	10.47	**6.746**	6.477
Maze	30	3.304	1.604			**3.114**	1.490	3.240	1.720	3.264	1.639
	25	3.904	1.895			**3.605**	1.760	3.884	2.055	3.940	1.996
	20	4.877	2.361			**4.477**	2.251	4.689	2.446	4.931	2.489
	15	4.844	4.101			**4.259**	4.270	7.153	7.168	6.896	7.019
	10	7.674	7.737			8.203	7.404	6.865	5.571	**6.008**	9.869
	8	9.963	8.524			**5.909**	6.165	9.437	9.746	9.926	10.24
	6	13.58	11.56			**10.30**	7.604	19.72	19.53	18.64	27.91
Grid	30	0.921	0.415			**0.881**	0.155	0.919	0.274	0.900	0.230
	25	1.064	0.478			1.034	0.286	1.016	0.231	**1.000**	0.234
	20	1.319	0.594			**1.218**	0.252	1.235	0.299	1.226	0.290
	15	1.279	3.114			1.649	1.335	1.892	1.312	**1.242**	0.961
	10	2.080	3.326			2.027	1.967	**1.753**	1.705	1.989	1.682
	8	2.514	1.826	2.110	1.936	2.020	2.087	2.604	1.743	**2.015**	1.998
	6	2.883	5.238	2.947	2.513	**1.843**	1.471	3.517	2.479	2.852	2.397

Table 1. Performance of the algorithms with a different size of M in each map

Apart from having a more suitable architecture for security purposes, in as many as 95% of cases one of the methods presented in this work improves *Cycle* algorithm. The only case where *Cycle* algorithm performed better was in *Strongly Connected* map with $M = 8$. Most notably, these results indicate that regardless of which map is used, in all cases, at least one of the methods presented in this work improves *MSP* algorithm. Taking into account that both *Cycle* and *MPS* algorithms use a centralized and explicit coordinator scheme, this improvement in performance is significant. Finally, in 80% of cases *MPS* algorithm does not work due to partitioning problems. Portugal and Rocha (2010) describe the reasons of these problems. It is worth noting that the proposed solution does not have these problems.

6. Conclusions

Several dynamic and distributed collaborative multi-robot approaches for security applications at critical facilities have been developed. Thus, a team of robots endowed with patrolling behaviors based on learning models from game theory as well as a thorough study of such models in the context of the patrolling problem has been presented. As shown in section 5, a significant improvement in performance was obtained by using the proposed methods with respect to *Cycle* and *MPS* algorithms. Moreover, the distributed characteristics of these models offer solutions with several advantages, such as scalability, modularity, and incremental expandability. Furthermore, the behavior of the robots patrolling that are using the techniques of this work is non-deterministic, which is suitable for security applications due to the fact that intelligent intruders can learn patrolling paths, and based on this information, perform attacks to the protected system. The evaluation to support this claim is not part of the scope of this work. However, results in (Sak and Wainer, 2008) demonstrate that system protection based on not static solutions is less susceptible to be attacked.

Despite the good performance achieved with the models implemented, there are significant remaining questions for future research. Firstly, interference among robots arises when more than one robot utilizes the same edge. In order to avoid this interference, it is necessary to evaluate if the edge selected is used or not by other robots. Future research consisting of studying the behavior of these methods including such aspect is necessary. Secondly, the metric of the patrolling simulator to evaluate the performance of the algorithms only includes the idleness of each node, however it does not take into account if one edge connected to such node (in the case that it has more than one) is used or not. Thus, an interesting future research consisting of evaluating the behavior of the algorithms including such a restriction because it allows to have a more secure system. Finally, even though the expected payoff matrix defined has achieved suitable results, new definition matrices should be explored.

7. Acknowledgements

This work has been supported by the Robotics and Cybernetics Research Group at Technique University of Madrid (Spain), and funded under the projects "ROTOS: Multi-Robot system for outdoor infrastructures protection", sponsored by Spain Ministry of Education and Science (DPI2010-17998), and the project ROBOCITY 2030 Project, sponsored by the Community of Madrid (S-0505/DPI/000235).

8. References

[1] Abate, F. (1996). The Oxford Dictionary and Thesaurus: The Ultimate Language Reference for American Readers, *Oxford University Press*.

[2] Aguirre, O., Llausas, R., Lucero, C., Taboada, H., Espiritu, J. and Kiekintveld, C. (2011). A multi-objective evolutionary algorithm for intelligent patrolling, *In Proceedings of the 41st International Conference on Computers & Industrial Engineering*.

[3] Ahmadi, M. and Stone, P. (2006). A multi-robot system for continuous area sweeping tasks, *In Proceedings of the IEEE International Conference on Robotics and Automation*.

[4] Almeida, A., Ramalho, G., Santana, H., Tedesco, P., Menezes, T., Corruble, V. and Chevaleyre, Y. (2004). Recent advances on multi-agent patrolling, *Advances in Artificial Intelligence,SBIA*, Springer.

[5] An, B., Kempe, D., Kiekintveld, C. and Shieh, E. (2012). Security Games with Limited Surveillance, *AAAI 2012*.

[6] Brown, G. (1951). Iterative solution of games by fictitious play, *Activity analysis of Production and allocation*.

[7] Camerer, C. (1999). Experience Weighted Attraction Learning in Normal Form Games, *Econometrica*.

[8] Chevaleyre, Y. (2004). Theoretical analysis of the multi-agent patrolling problem, *International Conference on Intelligent Agent Technology*, IEEE Computer Society.

[9] Elmaliach, Y., Agmon, N. and Kaminka, G. (2007). Multi-robot area patrol under frequency constraints, *Annals of Mathematics and Artificial Intelligence* .

[10] Everett, H. (2003). Robotic security systems, *Instrumentation & Measurement Magazine, IEEE*.

[11] Farinelli, A., Iocchi, L. and Nardi, D. (2004). Multi-robot systems: a classification focused on coordination, *Systems, Man, and Cybernetics, Part B: Cybernetics, IEEE Transactions on*.

[12] Fudenberg, D. (1998). *The theory of learning in games*, The MIT Press.

[13] Glad, A. and Buffet, O. (2009). Self-organization of patrolling-ant algorithms, *Proceedings of the 2009 Third IEEE International Conference on Self-Adaptive and Self-Organizing Systems*, IEEE Computer Society.

[14] Hwang, K. (2009). Cooperative patrol planning of multi-robot systems by a competitive auction system, *ICCAS-SICE*.

[15] Machado, A., Ramalho, G., Zucker, J. and Drogoul, A. (2003). Multi-agent patrolling: An empirical analysis of alternative architectures, *Multi-Agent-Based Simulation II*.

[16] Oates, R., Milford, M. and Wyeth, G. (2009). The implementation of a novel, bio inspired, robotic security system, *Intelligent Robots and Systems (IROS)*.

[17] Portugal, D. and Rocha, R. (2010). Msp algorithm: multi-robot patrolling based on territory allocation using balanced graph partitioning, *Proceedings of the 2010 ACM Symposium on Applied Computing*, ACM.

[18] Portugal, D. and Rocha, R. (2011). A survey on multi-robot patrolling algorithms, *Technological Innovation for Sustainability*.

[19] Sak, T. and Wainer, J. (2008). Probabilistic multi-agent patrolling, *Advances in Artificial Intelligence-SBIA*.

[20] Santana, H., Ramalho, G., Corruble, V. and Ratitch, B. (2004). Multi-agent patrolling with reinforcement learning, *Proceedings of the Third International Joint Conference on Autonomous Agents and Multi-agent Systems-Volume 3*, IEEE Computer Society.

[21] Sempé, F. and Drogoul, A. (2003). Adaptive patrol for a group of robots, *Intelligent Robots and Systems (IROS)*.

Space and Defense

Volume 12
Number 0 *Volume 12 Issue 1 (Winter 2021)*

Article 6

January 2021

Deterrence in Cyberspace: A Game-Theoretic Approach

Abderrahman Sokri
Center for Operational Research and Analysis, abderrahmane.Sokri@edu.edu

Follow this and additional works at: https://digitalcommons.unomaha.edu/spaceanddefense

Part of the Asian Studies Commons, Aviation and Space Education Commons, Defense and Security Studies Commons, Eastern European Studies Commons, International Relations Commons, Leadership Studies Commons, Near and Middle Eastern Studies Commons, Nuclear Engineering Commons, Science and Technology Studies Commons, and the Space Vehicles Commons

Please take our feedback survey at: https://unomaha.az1.qualtrics.com/jfe/form/SV_8cchtFmpDyGfBLE

Recommended Citation

Sokri, Abderrahman (2021) "Deterrence in Cyberspace: A Game-Theoretic Approach," *Space and Defense*: Vol. 12: No. 0, Article 6.
DOI: 10.32873/uno.dc.sd.12.01.1084
Available at: https://digitalcommons.unomaha.edu/spaceanddefense/vol12/iss0/6

This Article is brought to you for free and open access by DigitalCommons@UNO. It has been accepted for inclusion in Space and Defense by an authorized editor of DigitalCommons@UNO. For more information, please contact unodigitalcommons@unomaha.edu.

Deterrence in Cyberspace: A Game-Theoretic Approach

Cover Page Footnote

Abderrahmane Sokri is data scientist at Defence Research and Development Canada, Center for Operational Research and Analysis. Copyright for this article remains with the Crown, Canada.

This article is available in Space and Defense: https://digitalcommons.unomaha.edu/spaceanddefense/vol12/iss0/6

Deterrence in Cyberspace: A Game-Theoretic Approach

Abderrahmane Sokri

This novel application of the Stackelberg leader-follower game from economic theory illuminates situational constraints that point to a sweet spot, an optimal level of investment in cyber defense, for deterrence by denial.

Deterrence is a form of persuasion intended to manipulate the cost-benefit analysis of would-be attackers and convince them that the cost of taking an action against the defender outweighs its potential benefit (Brantly, 2018; Wilner, 2017).[1] It is the prevention (of a target) from committing unwanted behavior by fear of the consequences (United States (US) Department of Defense (DoD), 2008; Taipale, 2010). Deterrence differs from compellence by focusing on prevention using *ex ante* actions. Compellence uses power to force an adversary, *post hoc*, to take a desired action under threat of possible escalation in the future (Brantly, 2018).

Two types of deterrence are generally used: deterrence by punishment and deterrence by denial. Deterrence by punishment hinges on the threat of retaliation against a potential attacker. This tit-for-tat or equivalent retaliation strategy adds to the attacker's perceived cost. Deterrence by denial sends a signal to potential challengers that they will be unsuccessful. This impenetrability strategy subtracts from the attacker's perceived benefits.

In the physical world, deterrence aims to dissuade specific actions against physical assets. In this space, the most common form of deterrence by punishment is the use of nuclear weapons. These weapons are inherently an existential threat against potential challengers (Brodie et al., 1946; Brantly, 2018). An all-out nuclear war could be threatened but never fought to achieve reasonable political objectives (Freedman, 2004; Brantly, 2018). Deterrence by denial may include tightening defense around a critical infrastructure to deny attacker access. The target can be tightly defended by installing, for example, more security mechanisms and higher walls.

In the cyber domain, deterrence is more complex than in the physical domain. Digital attacks go beyond geographic and political boundaries. They are generally highly dynamic and imperceptible to the human senses (Moisan and Gonzalez, 2017; Sokri, 2019b). A cyber-attack may result in interception, degradation, modification, interruption, fabrication, or unauthorized use of an information asset. The information asset can be physically (e.g., hardware) or logically (e.g., software) based (Sokri, 2019a).

Cyber-attacks can be segregated into two main categories: targeted attack and opportunistic attack. A targeted attack requires a large effort and has the potential to cause significant damage to the defender. Denial of service and theft of information are typical targeted attacks. In contrast, an opportunistic attack has a number of intermediate targets, requires a small effort, and tends to cause less damage. A virus and spam e-mail are typical opportunistic attacks.

---

[1] Abderrahmane Sokri is data scientist at Defence Research and Development Canada, Center for Operational Research and Analysis. Copyright for this article remains with the Crown, Canada.

The most challenging problem in cyber deterrence is the attribution dilemma (Wilner, 2017). Determining who to blame for an attack may be very difficult and time-consuming to do. Consequently, the credibility of any deterrence by punishment in digital space will depend on the blame attribution. (Glaser, 2011; Brantly, 2018). Since deterrence by denial does not require identification of potential attackers, it can be used to mitigate this dependency (Bordelon, 2016).

Cyber risk is present when a given threat meets a vulnerability in an information system allowing it to manifest. In this context, a threat is a potential cause of an unwanted occurrence while a vulnerability is a weakness in the information system (Sokri, 2019a; Zhang, 2012; Bowen et al., 2006). To minimize digital risk against an information asset, the defender should know at least two elements: (1) the probability of a successful attack and (2) the corresponding potential loss (Brantly, 2018; Glaser, 2011; Schneidewind, 2011; Branagan, 2012).

To protect their information assets against offensive cyber-attacks, policy makers are increasingly gravitating towards deterrence by denial (Taipale, 2010). A key decision-variable in digital deterrence by denial is the defender investment level in security. To protect a potential target, the defender can reduce the probability of a successful attack by investing in information security. The investment may, for example, reduce the vulnerability of the target.

The aim of this paper is to show how deterrence by denial as a defense strategy can be formulated in cyberspace using a sequential game with a disclosure mechanism. It shows the suitability of game theory to cyber deterrence. The paper extends existing models by providing a new game formulation of deterrence using a more intuitive probability of a successful attack. It also combines stochastic simulation and game-theoretic approaches to handle uncertainty in the input data. A simulation could, for example, incorporate uncertainty on the model variables and parameters by changing their static values to statistical distributions.

Consider a sequential security game played between two adversarial agents: a defender *D* (the leader) and a strategic attacker *A* (the follower). The defender anticipates the attacker's reaction, determines, and credibly communicates the security investment to protect an information system. The defender can, for example, publicly release his level of investment in (1) detection and prevention techniques such as Antivirus software, Firewalls, and Intrusion Detection Systems (IDS) and (2) physical monitoring and inspection procedures (Sokri, 2019b). Revenue agencies usually use this tactic by revealing their auditing strategies to deter tax evasion (Cavusoglu et al., 2008).

The attacker observes the defender's decision and reacts with a certain level of willingness-to-attack. The true willingness-to-attack is latent and, therefore, not directly observable. It is modeled as the expected effort to be exerted by the attacker to compromise the system. The attacker's effort corresponds to the first activities of the cyber kill chain (Mihai et al., 2014). These activities particularly include (but are not limited to):

1. **Reconnaissance** – the process of collecting information about the system,
2. **Weaponization** – the process of analyzing the collected data to select the appropriate attack technique, and
3. **Delivery** – the process of transmitting the weapon to the targeted system.

Following this introduction, section 2, below, provides a comprehensive review of literature on security investment as a deterrence factor. Section 3, sets up a new game theoretic model of deterrence in cyberspace. Section 4, computes the Stackelberg equilibrium. Section 5 offers a formal discussion about the main results. Some concluding remarks are indicated in section 6.

LITERATURE REVIEW

Identifying and understanding the factors influencing the decision to invest in information security is a key requirement for any effective deterrence and risk management in cyberspace. These factors form the pillars of the appropriate level of security investment. Security investment as a deterrence factor has been an active research area in the last decade. This literature can be divided into two main categories: decision theory and game theory approaches (Cavusoglu et al., 2008).

The decision-theoretic approach uses traditional risk analysis and cost–benefit perspectives for security investment decisions. This approach assesses the risk associated with security breaches and conducts a cost-benefit analysis to determine a certain level of security investment to mitigate the risk. While this approach can assess the economic value of intangible costs and benefits, it has two main limitations: (1) It does not determine the optimal security investment level. (2) It does not allow a defender's security investment to influence the attacker's behaviour.

Al-Humaigani and Dunn (2003), for example, proposed a model to quantify the return on security investment (ROSI). The authors enumerated the fundamental components of ROSI for every organization and security threat. They included what it costs to invest in information security spending (e.g., the cost of procuring the security tool or software, the losses in reputation and goodwill). They incorporated both the pre- and post- system implementation security measures.

In order to come through the first limitation of the decision-theoretic approach, Gordon and Loeb (2002) presented an economic model that determines the optimal amount to invest in information security. Their results indicate that defenders may be better off concentrating their efforts on information assets with midrange vulnerabilities. Extremely vulnerable information assets may be very expensive to protect. For some broad classes of security breach probability functions, results also indicate that optimal investment never exceeds 37% of the expected loss. Hausken (2006) examined the effect of different returns assumptions on the optimal level of investment. The author showed that optimal investment level may no longer be capped at 37% of expected loss. For an alternative class of security breach probability functions, the optimal investment can increase convexly in vulnerability and exceed 37%.

More recently, Mayadunne and Park (2016) used the expected utility approach to analyze information security investment decisions. They provided a comparison between the decisions made by a risk taking and a risk neutral decision maker. They found, for example, that for a group of information assets with equal value and varying vulnerabilities, the risk neutral decision maker will diversify security investment to a greater extent and the risk taker will invest a larger amount when protecting the high risk assets in the group.

The game-theoretical approach uses game oriented models to capture the strategic interactions between rational attackers and

defenders. Optimal investment in security is one of the defenders' resulting strategies. This approach has two main challenges: (1) Validity of the game-theoretic assumptions in cyberspace (e.g., rationality of players). (2) Complexity of the cyber domain scenarios (e.g., dynamic attacks and complex networks).

Cavusoglu et al. (2008), for example, argued that the old decision-theoretic approach is incomplete because it does not take into account the strategic nature of the interaction between attackers and defenders. The authors used a game-theoretic model to determine the optimal security investment level. Results indicate that the defender generally enjoys a higher payoff than that in the decision theory approach. The gap between the two results decreases over time and the rate of convergence depends on the defender learning model.

Wu et al. (2015) used game theory to model the relationship between the optimal information security investment and the characteristics of defenders' security environment. Results indicate that defenders are better off not investing in security (outside best practices) until the potential loss reaches a certain value. They should focus on the midrange of intrinsic vulnerabilities. When the potential loss is catastrophic, they should adopt other measures and stop investing in security.

More recently, Pan et al. (2017) suggested an optimal investment strategy using a game-theoretic framework. The authors concluded that the defender is better off using a single security level to protect all the information assets instead of using different security levels to protect different assets. The interested reader is referred to Sokri (2019a) and Sokri (2019b) for further information on game theory in cyber defense.

A "STACKELBERG" DETERRENCE MODEL

The system is characterized by an inherent vulnerability v_0. Each successful attack can result in a potential loss l to the defender and a possible benefit b to the attacker. The loss/benefit occurring can be tangible (e.g., monetary loss/benefit) or intangible (e.g., loss/gain in reputation).

Probability of a successful attack

Let i be the defender's security investment and t the attacker's level of effort to expend in hacking the defender. The compound probability p of a successful attack can be expressed as the product of the probability that the vulnerability may be exploited, $v(i)$, and the threat probability (i.e., the probability to receive an attack) (Wu et al, 2015):

$$(1) \quad p = v(i)\left(1 - exp\left(-\frac{t}{\mu}\right)\right),$$

where the expected effort t can be expressed in terms of time. The threat probability, also known as the probability of attack (prior to information about target vulnerability), is written in Equation 1 as the cumulative distribution function (CDF) of an exponentially distributed random variable evaluated at t. This CDF estimates the probability that the attacker's level of effort will be less than t. The parameter μ represents the mean effort to attack (e.g., investigation, identification, weaponization done prior to knowledge of target defenses). It also represents the standard deviation of the distribution.

As in Wu et al. (2015), the defender's security investment does not directly affect the inherent threat probability. The defender can only reduce the first term, probability that the vulnerability may be exploited, using security investment i. That is,

(2) $v(i) = v_0 \exp(-\alpha i)$,

where the parameter $\alpha > 0$. Straightforward derivation leads to

(3) $\frac{v'(i)}{v(i)} = -\alpha$,

which means that the parameter α is the decay rate of the probability that the vulnerability may be exploited. It represents the rate at which vulnerability decreases with investment in cybersecurity. It can also be seen as a measure of investment productivity. It measures how efficiently security investment is used to reduce the asset vulnerability.

One can also readily see that $v(i)$ satisfies the following three assumptions.

- Assumption 1. $v(0) = v_0$.
- Assumption 2. $\lim_{i \to \infty} v(i) = 0$.
- Assumption 3. $v'(i) = \frac{dv(i)}{di} < 0$, $v''(i) = \frac{d^2 v(i)}{di^2} > 0, \forall i$.

Assumption 1 states that if there is no investment in security, the vulnerability of the system will be the inherent vulnerability. Assumption 2 states that no finite investment can eradicate the vulnerability from information systems. Because of their complexity, perfect security is impossible (Wu et al., 2015). Assumption 3 states that the investment in security reduces the probability that the vulnerability may be exploited, but at a decreasing rate. Investment makes the system more secure, but with declining marginal return.

The probability of vulnerability exploitation is formulated in Equation 2 as an exponentially decreasing function of the security investment. Consequently, the probability of a *successful* attack can now be written as

(4)

$$p(i,t) = v_0 \exp(-\alpha i)\left(1 - \exp\left(-\frac{t}{\mu}\right)\right).$$

This probability depends on the defender investment level and the attacker's effort level, in addition to the system's inherent vulnerability.

Defender's loss and attacker's payoff

In this game the defender seeks to find the optimal security investment that minimizes the following total cost

(5) $W_D = p(i,t)l + i$,

where the first term of its right-hand side is the defender's expected loss due to a successful attack. The attacker seeks to maximize the following payoff

(6) $W_A = p(i,t)b - t$,

where the first term of the right-hand side is the attacker's expected benefit and the second term represents the expected effort to compromise the system.

Deterrence game's equilibrium

This section characterizes the optimal solution to the deterrence game. As in the standard Stackelberg competition, the game is sequential: the defender moves first, committing to a strategy before the attacker reacts. The defender's strategic choice is to select the optimal security investment (deterrence by denial). The attacker's choice is to determine his appropriate level of effort. The outcome of this leader-follower interaction is called Stackelberg equilibrium. This equilibrium has been recognized as a sound theoretical framework for modeling the strategic interactions between attackers and defenders (Jain et al., 2010; Korzhyk et al., 2011; Kiekintveld et al., 2015; Acquaviva, 2017).

Proposition 1. The following condition is satisfied at equilibrium

(7) $\frac{\partial p(i,t)}{\partial t} b = -\frac{\partial p(i,t)}{\partial i} l.$

Proof. Assuming an interior solution, the first-order condition (maximizing attacker payoff with respect to effort, t) for the attacker optimization problem is

(8) $\frac{dW_A}{dt} = \frac{\partial p(i,t)}{\partial t} b - 1 = 0.$

The optimality condition for the defender problem is

(9) $\frac{dW_D}{di} = \frac{\partial p(i,t)}{\partial i} l + 1 = 0.$

Equations 8 and 9 lead to the equilibrium condition in the Proposition.

■

Fixing the defender's security investment to some strategy i, the first problem to be solved is to find the attacker's best response to i. In this optimization problem, the follower maximizes his expected benefit given i.

Proposition 2. Assuming an interior solution, the optimal effort the attacker is willing to exert is given by

(10) $t = -\alpha\mu i + \mu \ln\left(\frac{bv_0}{\mu}\right)$

Proof. After substitution for $p(i,t)$, Equation 6 becomes

(11)
$$W_A = v_0 \exp(-\alpha i)\left(1 - \exp\left(-\frac{t}{\mu}\right)\right) b - t.$$

Computing the derivative of W_A with respect to t, equating to zero, and solving leads to the expression of t as a function of i.

■

Proposition 3. The attacker's level of effort is a decreasing function in the defender's investment.

Proof. The derivative of t with respect to i is

(12) $\frac{dt}{di} = -\alpha\mu < 0.$

■

Proposition 4. Assuming an interior solution, the defender optimal security investment level is given by

(13) $i = \frac{1}{\alpha} \ln(\alpha l v_0).$

Proof. Equations 4 and 5 imply that

(14)
$$W_D = v_0 \exp(-\alpha i)\left(1 - \exp\left(-\frac{t}{\mu}\right)\right) l + i.$$

The expression of t in Equation (10) is equivalent to

(15) $\exp\left(-\frac{t}{\mu}\right) = \frac{\mu}{bv_0} \exp(\alpha i).$

Substituting for $\exp\left(-\frac{t}{\mu}\right)$ from Equation 15 in Equation 14, computing the derivative of W_D with respect to i, equating to zero, and solving provides the equilibrium strategy in the Proposition.

■

Proposition 5. The attacker's optimal level of effort is given by

(16) $t = \mu \ln\left(\frac{b}{\alpha\mu l}\right).$

Proof. Substituting for i from Equation 13 in Equation 10 leads to the result.

■

Proposition 6. The defender should not invest in security beyond best practices until the potential loss reaches

(17) $l^* = \frac{1}{\alpha v_0}.$

Proof. To have a positive investment, $\ln(\alpha l v_0) > 0$. This is possible only if $\alpha l v_0 > 1$, which leads to the condition in the Proposition.

■

Proposition 7. The attacker should not exert any effort until the potential benefit reaches

(18) $b^* = \alpha \mu l.$

Proof. To have a positive effort, $\ln\left(\frac{b}{\alpha \mu l}\right) > 0$. This is possible only if $\frac{b}{\alpha \mu l} > 1$ which leads to the condition in the Proposition.

■

Proposition 8. The defender's optimal security investment level is an increasing concave function of the potential loss, l.

Proof. As shown in Equations 19 and 20, the first derivative of i with respect to l is positive and the second derivative is negative, respectively.

(19) $i'(l) = \frac{di(l)}{dl} = \frac{1}{\alpha l} > 0.$

(20) $i''(l) = \frac{d^2 i(l)}{dl^2} = -\frac{1}{\alpha l^2} < 0.$

Consequently, i is a concave function in l that increases at decreasing rate.

■

Proposition 9. The defender's optimal investment spent on information security as a fraction of potential loss l is given by

(21) $r(l) = \frac{i}{l} = \frac{1}{\alpha l}\ln(\alpha l v_0).$

Proof. Dividing the expression of i in Equation 13 by l leads to the result.

Proposition 10. The fraction $r(l)$ is an increasing function in the potential loss l for $l \leq l^{**} = \frac{e}{\alpha v_0} \approx \frac{2.718}{\alpha v_0}$. It is decreasing for $l \geq l^{**}$ with a horizontal asymptote at $y = 0$.

Proof. The first derivative of $r(l)$ with respect to l is

(22) $r'(l) = \frac{dr(l)}{dl} = \frac{1}{\alpha l^2}\left(1 - \ln(\alpha l v_0)\right).$

It is straightforward to show that $r'(l) = 0$ for $l = \frac{e}{\alpha v_0}$, $r'(l) \geq 0$, for $l \leq \frac{e}{\alpha v_0}$, and $r'(l) \leq 0$, for $l \geq \frac{e}{\alpha v_0}$.
Hence, the potential loss l that maximizes the fraction $r(l)$ is given by

(23) $l^{**} = \frac{e}{\alpha v_0} \approx \frac{2.718}{\alpha v_0}.$

Using the l'Hopital rule, $\lim_{l \to \infty} r(l) = 0$, which shows that $r(l)$ has a horizontal asymptote at $y = 0$.

■

To deal with uncertainty in the input data, a Monte Carlo simulation could represent each uncertain parameter as a probability distribution.

INVESTMENT IN CYBERSECURITY AT EQUILIBRIUM

A parsimonious game-theoretical model is used in this paper to characterize deterrence in cyberspace. A Stackelberg game is played to capture the strategic nature of this interaction and provide clear insights about it. The suggested mechanism involves disclosing the defender's investment information to the potential attacker. The game's logic and results crucially depend on the timings of each move. The defender moves first, anticipates the strategic behavior of the attacker, and decides on the security investment. The attacker observes the defender's level of investment and determines a certain effort level. By revealing the security investment strategy, the defender becomes able to control the attacker's

incentive and deter (or reduce the effort behind) potential attacks.

Assuming an interior solution, Proposition 1 characterizes the first-order optimality conditions for the defender and attacker strategies. It compares, at equilibrium, magnitude decline in expected defender loss from extra security investment to magnitude increase in expected attacker benefit from extra effort. Stackelberg interaction joins their fates.

At equilibrium, marginal reduction in defender's expected loss due to additional investment precisely balances marginal increase in the attacker's expected benefit attributable to additional effort. In order to reach this decision point, the attacker as follower must be able to measure the magnitude of loss to the defender from a successful cyber attack. In the Stackelberg interaction, attacker does have a clue from observing optimal defender security investment, which is tied to defender assessment of cost in the event of disruption. Physical properties of the cyber system's vulnerability must also be common knowledge.

Propositions 4 and 5 define the attacker's optimal level of effort and the defender's optimal investment, respectively. Proposition 4 relates the defender's strategy to three parameters:

- the inherent vulnerability v_0
- the decay rate in the vulnerability due to investment α, and
- the defender's potential loss l.

Proposition 5 shows that the attacker's strategy depends on two other parameters in addition to α and l, namely the mean level of effort μ (independent of system vulnerability) and the attacker potential benefit b from system disruption.

The derivative of the attacker's expected effort t with respect to the defender's investment i in Equation 12 indicates that the parameters α and μ and their interaction effect are the key factors in cyber deterrence, that is, in sharply affecting adversaries' attack plans through denial. Equation 12 shows that the higher the two parameters the more likely the attacker is to be deterred through additional defender investment. The parameter α measures the speed at which security investment translates into a reduction of the asset's vulnerability to attacks. An increase in the parameter α for any given level of investment will decrease the probability that inherent vulnerability may be exploited, lessen the probability of a successful attack, and, therefore, result in a reduction in the attacker's level of effort. At the same time, the influence of additional investment on reducing attacker effort even further will rise. Equation 12 also shows that opportunistic attacks (with small μ) are harder to influence than targeted attacks (with high μ). Extensive initial interest in the targeted system leads potential attackers to be discouraged at a steeper rate once they learn of additional defender investment.

Propositions 6, 8 and 9 characterize the defender's optimal security investment level i as a function of the potential loss l. These propositions highlight the following key findings:

- The defender should not invest in security beyond best practices until the potential loss reaches a given value;
- The optimal security investment increases with the expected loss at a decreasing rate;
- The optimal investment in security as a fraction of potential loss l has a horizontal asymptote at $y = 0$. This means that, for very large potential losses, the optimal amount to spend on information security does not keep

pace; it is far smaller than the potential loss.

These findings are on par with the deterrence literature. They are particularly consistent with the study conducted by Gordon and Loeb (2002).

The formalism in Equations 4, 5, and 6 is grounded theoretically such that the model could be repeated or extended using different probability distributions. Its underlying mathematics is clear and conceptually based. Variations of the probability distribution will provide qualitatively the same findings. The numerical values of these findings will, of course, depend on the values of the deterrence model parameters.

A myopic approach such as a simultaneous game or a decision-theoretic technique would produce different results. Under a simultaneous game, players make single decisions before seeing the other player's moves (as in the famous Prisoner's Dilemma [PD]) and possibly under incomplete information about the other player's payoff from certain outcomes. Attackers, for example, are not able to observe the outcome of previous actions before responding. The main characteristic of myopic approaches is the non-cooperative, monotonic relationship between defender investment level and attacker effort. Both players rationally defect in PD-type games. When one cost variable increases, the other increases and vice versa; net payoffs in equilibrium for both decline. In this situation, attackers are never deterred, *per se*, because myopic approaches lack disclosure mechanisms. A deeper understanding of this interaction will be generated in future works.

CONCLUSION

Deterrence is used to prevent unwanted actions by influencing the cost-benefit analysis of potential attackers. The most common form of deterrence in cyberspace is deterrence by denial. Deterrence by denial sends a signal to would-be attackers that they will be unsuccessful. In this defense strategy, the defender reduces the probability of a successful attack by investing in information security. While the credibility of deterrence by punishment depends on blame attribution, deterrence by denial does not require this knowledge.

This paper used a sequential game theoretic approach with a disclosure mechanism (Stackelberg competition) to formulate a deterrence strategy in cyberspace. It derived the defender's optimal security investment level and the attacker's level of effort. The factors influencing the decision to invest in cybersecurity were identified and discussed. To deal with uncertainty in the input data, the model invites parametric analysis using Monte Carlo simulation.

Results for the equilibrium indicate that effectiveness of the security investment (α) and the category of attack (μ) and their interaction effect are the key factors in cyber deterrence. The more effective the security investment in reducing vulnerability and the higher attacker initial interest in the target, the more likely attacker is to be deterred by additional investment. Targeted attacks aiming at significant damage to the defender are more manageable by security investment than opportunistic attacks.

The defender's optimal security investment level (i) as a function of potential loss (l) indicates that investment in cybersecurity as a deterrence strategy will top out after the middle part of losses. At very high levels of loss, there is a numbing effect; optimal investment does not change much with additional increments of loss.

Deterrence in the cyber domain is more complex than in the physical field. Further efforts should be undertaken to understand it in order to influence potential attackers' behaviors. Examples of such studies include (but are not limited to)

- application of the model to a real-world cyber-security problem using real-life parameters,
- analyzing the interaction between defenders and attackers in dynamic scenarios,
- assessing the risk to the defender of a disclosure strategy,
- including deception mechanisms to enhance security,
- developing models to deal with bounded rationality of human adversaries,
- combining game theoretic models such as this Stackelberg version with other techniques and tools to make the formalism more realistic and tractable; techniques may include numerical simulation and genetic algorithms; tools may consist of firewalls and anti-virus software.

REFERENCES

Al-Humaigani, M., and Dunn, D. (2003). A model of return on investment for information systems security. IEEE 46th Midwest Symposium on Circuits and Systems, Vol. 1, pp. 483-485.

Acquaviva JR (2017). Optimal Cyber-Defence Strategies for Advanced Persistent Threats: A Game Theoretical Analysis. Master Thesis, the Pennsylvania State University.

Bordelon, E.B. (2017). Approaching Cyber Warfare: Geopolitics, Deterrence, and International Law. Senior Honors Thesis 639, Cyber War and Geopolitcs, Liberty University, Virginia.

Bowen, P., Hash, J., and Wilson, M. (2006). *Information Security Handbook: A Guide for Managers*. National Institute of Standards and Technology (NIST) Special Publication 800-100.

Branagan, M. (2012) A risk simulation framework for information infrastructure protection. Ph.D. Dissertation, Queensland University of Technology, Australia.

Brantly, A.F. (2018). The cyber deterrence problem. 10th International Conference on Cyber Conflict (CyCon), IEEE, pp. 31-54.

Cavusoglu, H., Raghunathan, S., and Yue, W.T. (2008). Decision-Theoretic and Game-Theoretic Approaches to IT Security Investment. *Journal of Management Information Systems*, Vol. 25:2, pp. 281-304.

Freedman, L. (2004). *Deterrence*. Polity Press, Cambridge.

Glaser, C.L. (2011). Deterrence of Cyber-attacks and US National Security. Report GW-CSPRI-2011-5, The George Washington University, Washington, D.C.

Gordon, L. A., and Loeb, M. P. (2002). The economics of information security investment. *ACM Transactions on Information and System Security* (TISSEC), Vol. 5:4, pp. 438-457.

Hausken, K. (2006). Returns to information security investment: The effect of alternative information security breach functions on optimal investment and sensitivity to vulnerability. *Information Systems Frontiers*, Vol. 8:5, pp. 338-349.

Jain M, Tsai J, Pita J, Kiekintveld C, Rathi S, Ordone, F, and Tambe, M. (2010). Software assistants for randomized patrol planning for the LAX airport police and the Federal Air Marshals Service. *Interfaces*, Vol. 40:4.

Kiekintveld C, Lisy V, and Pibil R. (2015). Game-theoretic foundations for the strategic use of honeypots in network security. In *Cyber Warfare*. Springer, p. 81–101.

Korzhyk D, Yin Z, Kiekintveld C, Conitzer V, and Tambe M. (2011). Stackelberg vs. Nash in

Security Games: An Extended Investigation of Interchangeability, Equivalence, and Uniqueness. *Journal of Artificial Intelligence Research*, Vol. 41.

Mayadunne, S., and Park, S. (2016). An economic model to evaluate information security investment of risk-taking small and medium enterprises. *International Journal of Production Economics*, Vol. 182, pp. 519-530.

Mihai, I.C., Pruna, S. and Barbu, I.D. (2014). Cyber Kill Chain Analysis. *Int'l J. Info. Sec. & Cybercrime*, Vol. 3.

Moisan, F. and Gonzalpez, C. (2017). Security under Uncertainty: Adaptive Attackers Are More Challenging to Human Defenders than Random Attackers. *Frontiers in Psychology*, Vol. 8:982.

Pan, C., Zhong, W., and Mei, S. E. (2017). Investment strategy analysis of information systems with different security levels. IEEE 2nd International Conference on Big Data Analysis, pp. 703-708.

Pereira, J. P., & Ferreira, P. (2011). Next Generation Access Networks (NGANs) and the geographical segmentation of markets. ICN Tenth International Conference on Networks, pp. 69-74.

Schneidewind, N.F. (2009). *Systems and Software Engineering with Applications*. Wiley-IEEE Press.

Sokri, A. (2019a). Cyber Security Risk Modelling and Assessment: A Quantitative Approach. Proceedings of the 19th European Conference on Cyberwarfare and Security (ECCWS19), 4-5 July 2019, Coimbra University, Portugal.

Sokri, A. (2019b). Game theory and cyber defence. In: *Games in Management Sciences*, Pineau, P.-O. and Taboubi, S. (eds) Springer International Series in Operations Research & Management Science.

Taipale, K.A. (2010). Cyber-deterrence. Law, Policy and Technology: Cyberterrorism, Information, Warfare, Digital and Internet Immobilization, IGI Global.

US DoD (2008). Joint Publication 1-02, Department of Defense Dictionary of Military and Associated Terms, Washington, D.C.

Wilner, A. (2017). Cyber deterrence and critical-infrastructure protection: Expectation, application, and limitation. *Comparative Strategy*, Vol. 36:4, pp. 309-318.

Wu, Y., Feng, G., Wang, N., and Liang, H. (2015). Game of information security investment: Impact of attack types and network vulnerability. *Expert Systems with Applications*, Vol. 42:15-16, pp. 6132-6146.

Zhang, J. (2012). Information Security Risk Management Framework: China Aerospace Systems Engineering Corporation. Master Thesis, University of South Australia.

CC BY-NC-SA 4.0
https://doi.org/10.4467/27204383TER.23.029.18331
https://orcid.org/0000-0002-5288-0324
https://orcid.org/0000-0002-3907-2242
https://orcid.org/0000-0001-6390-1402

TOMASZ P. MICHALAK
MICHAŁ T. GODZISZEWSKI
ANDRZEJ NAGÓRKO

Protecting critical infrastructure with game theory, optimization techniques, and AI algorithms

Abstract

In light of recent geopolitical developments, Europe and Poland are acutely aware of the urgent importance of infrastructure security. Despite heightened interest and increased investments, our security resources remain severely limited, rendering continuous protection for every potential target unattainable. Consequently, the strategic allocation of security resources becomes an ongoing imperative. This paper presents a short introduction to the core principles behind advanced methods that facilitate automated decision-making in security resource allocation. These methods leverage artificial intelligence (AI), game theory, and optimization techniques, and have demonstrated their effectiveness through multiple real-life deployments in the USA. We also provide a concise overview of this exciting body of research and discuss the solutions and software developed by our team, "AI for Security" at the IDEAS NCBR research institute to protect critical infrastructure in Poland and in Europe.

Keywords:

optimization, security games, artificial intelligence, critical infrastructure

Introduction

Los Angeles International Airport (LAX) is one of the busiest and largest airports in the world, serving as a vital transportation hub for the city of Los Angeles and the surrounding region. In terms of the number of passengers, it is about four times larger than the Frederic Chopin Airport, Warsaw, which is the biggest Polish airport. The LAX airport encompasses a vast area and features four parallel runways and nine terminals each serving different airlines and destinations. The largest one is the Tom Bradley International Terminal dedicated to international flights. The Central Terminal Area serves as the focal point of the airport connecting all the terminals. It includes a complex network of roadways, parking structures, and transportation services, such as shuttles and taxis, to facilitate passenger movement around the airport.

Given its prominence and size, LAX is one of the prime targets on the West Coast of potential attacks. Safeguarding such a complex and sprawling facility requires a delicate balance between security measures and operational efficiency. Unfortunately, available security resources are limited, making it impossible to provide round-the-clock security for each and every place of interest. For instance, while the number of canines exceeds the number of terminals, there are only a handful of canines of particular expertise, such as explosive detection ones. It means that it is simply impossible to provide constant coverage of a single explosive detection canine patrol per each terminal at LAX. The same hold for other types of canine patrols, such as those specialised in drug detection.

Interestingly, however, in 2008, the drug sniffing canine patrols at the LAX airport turned out to be much more effective than it was previously believed possible. While in the 15 month period between April 2006 to July 2007, only 4 drug-related offences were recorded, in the 12 months of 2008 the number of cases grew to 30. The reason behind this significant improvement was ARMOR which stands for the Assistant for Randomized Monitoring over Routes. The ARMOR system was an innovative software tool developed by Milind Tambe and his colleagues at the University of Southern California (USC)[1], within the Department of Homeland Security's first University Center of Excellence.

[1] J. Pita et al., *Using game theory for Los Angeles Airport security*, "AI Magazine" 2009, vol. 30, no. 1, pp. 43–57.

The primary objective of the ARMOR system is to assist security personnel in making better and more efficient decisions by weighing risks against available resources. To this end, ARMOR leverages artificial intelligence (AI), game theory, and optimization techniques. The system makes it difficult for adversaries to plan how to avoid security forces during an attack. Even more importantly, it allows security forces to deploy their limited resources in the most efficient manner and achieve maximal effectiveness.

The implementation of ARMOR at LAX resulted in the improvement in security coverage, resource allocation efficiency, and deterrence against potential attackers. For example, Chart shows that the introduction of the ARMOR system at the LAX Los Angeles Airport resulted in more than threefold increase in the number of detected offences. The system serves as an exemplar of how advanced technology and AI-driven approaches can contribute to strengthening security measures and enhancing the safety of airports and their passengers.

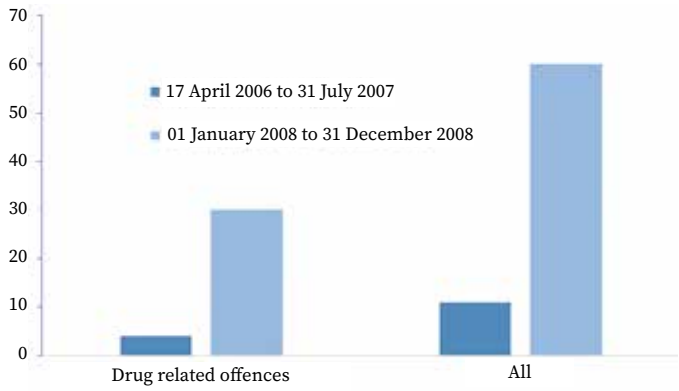


Chart. The number of offences in the period of 15 months before introducing ARMOR (dark blue bars) vs. the number of offences in the period of 12 months after introducing the system (light blue bars).

Source: own elaboration based on: Pita et al., *Using game theory for Los Angeles Airport security*, "AI Magazine" 2009, vol. 30, no. 1, pp. 43–57.

The success of ARMOR received significant attention in the context of security applications. A number of systems based on similar principles were deployed in the USA to protect other critical infrastructure sites. These include, in particular:

- IRIS system[2] – to optimize the routes and schedule of the security agents in the U.S. Air Marshals program;
- PROTECT[3] – to optimize the security of the Boston and New York ports;
- TRUSTS system[4] – to prevent fare evasion created for the railway transport system in Los Angeles.

Furthermore, it has been advocated in the context of cybersecurity[5]. There are also a growing number of civilian applications such as protecting endangered species in national parks (systems PAWS[6] and MIDAS[7]). In all these cases, it was possible to significantly improve security, not by adding many additional security resources but by better deployment of the available ones.

This is a very important lesson for Europe and Poland in particular. Given recent geopolitical developments, and the on-going Russian full-scale invasion in Ukraine that started in 2022, we are all well aware of the pressing concern of infrastructure security. Europe has already witnessed a few such attacks[8]. The question now is not if the next attacks will happen but when.

---

[2] J. Tsai et al., *Iris - a tool for strategic security allocation in transportation networks*, in: *Proceedings of the 8th International Conference on Autonomous Agents and Multi-Agent Systems* (AAMAS 2009), pp. 37–44 (the proceedings of the AAMAS conference series are available at: https://dl.acm.org/conference/aamas/proceedings – editor's note).

[3] E. Shieh et al., *Protect: A deployed game theoretic system to protect the ports of the United States*, in: *Proceedings of the 11th International Conference on Autonomous Agents and Multi-Agent Systems* (AAMAS 2012), vol. 1, pp. 13-20.

[4] Z. Yin et al., *Trusts: Scheduling randomized patrols for fare inspection in transit systems*, in: *Proceedings of the Twenty-Sixth AAAI Conference on Artificial Intelligence* (AAAI 2012), vol. 26, no. 2, pp. 2348–2355 (the proceedings from the AAAI conferences and symposia are available at: https://aaai.org/aaai-publications/aaai-conference-proceedings/ – editor's note).

[5] Y. Zhang, P. Malacaria, *Bayesian Stackelberg games for cyber-security decision support*, "Decision Support Systems" 2021, vol. 148, art. 113599. https://doi.org/10.1016/j.dss.2021.113599.

[6] R. Yang et al., *Adaptive resource allocation for wildlife protection against illegal poachers*, in: *Proceedings of the 13th International Conference on Autonomous Agents and Multiagent Systems* (AAMAS 2014), pp. 453–460.

[7] W. Haskell et al., *Robust protection of fisheries with COmPASS*, in: *Proceedings of the Twenty-Eighth AAAI Conference on Artificial Intelligence* (AAAI 2014), vol. 28, no. 2, pp. 2978-2983.

[8] An example is the deliberate cutting of two optical fibers of the Deutsche Bahn communication system on October 8, 2022, which stopped rail traffic in northern Germany for approximately 3 hours.

Unfortunately, the problem is exacerbated by the growing technological sophistication of critical infrastructure. While modern communication, computation, and control technologies enhance the efficiency, they also make modern critical infrastructure systems increasingly complex and vulnerable to deliberate attacks and random failures. These attacks can manifest in various forms, severities, and magnitudes, ranging from acts of terrorism on local infrastructure to major kinetic strikes during times of war, such as the ongoing Russian invasion on Ukraine. New technologies, such as drones, also enhance the capabilities of the potential attackers.

Facing the evolving and expanding landscape of threats, despite increased interest and investments in infrastructure security, our security resources will remain limited, making it impossible to provide constant protection for everything. Thus, the need to strategically allocate security resources becomes a perpetual necessity. The example of the LAX airport as well as other aforementioned examples from the USA show that such well-designed strategic decision making is beneficial and delivers significant improvement in security.

In this paper, we discuss the fundamentals behind these advanced methods of protecting critical infrastructure, briefly review this body of research, and present the solutions and software which is developed by our team "AI for security" at the IDEAS NCBR research institute.

Defender-Attacker Security Games

Game theory studies interactions between intelligent entities like individuals, companies, or countries. In the context of security, these entities can represent "the defenders", e.g., security forces, police, military, and "the attackers", e.g., criminals, terrorists, and state actors. Game-theoretic approaches help us understand how these intelligent actors interact, assuming they can anticipate and respond to each other's actions. By using game theory, we can develop a strategy to efficiently distribute scarce security resources for infrastructure protection. This approach considers the importance of different targets and how adversaries may react to specific protection strategies.

A non-cooperative game is defined by the set of players, the set of strategies for each player and the payoff function that assigns each

player the utility for any possible combination of strategies. Each game is also associated with some set of rules, e.g., we may require the players to move simultaneously or sequentially.

Table 1 presents a sample game from Pita et al.[9] In this case, we have two players, each having two strategies: {*A*, *B*} and {*C*, *D*}, respectively. The values of the payoff function are given by the pairs of numbers in the matrix, where each cell corresponds to a given combination of strategies, also known as a 'strategy profile'. For instance, if Player 1 plays strategy and Player 2 plays strategy then the Player 1 earns the payoff of 2 while Player 2 the payoff of 1.

Sometimes it is possible to stipulate how the rational players (where the notion of rationality is explicated with a precise mathematical formula) actually would play the game given its rules. Such a collection of players' strategies is called an equilibrium of the game. Perhaps the most well-known equilibrium concept is the Nash equilibrium. A combination of strategies is a Nash equilibrium if any player would not like to change their strategy, given the strategies chosen by the opponents. For instance, in Table 1, the combination of strategies (*A*, *B*) is not a Nash equilibrium, because Player 2 would like to change their strategy from D to C, assuming that Player 1 sticks to strategy A. Conversely, the combination of strategies (*A*, *C*) is a Nash equilibrium, because, for Player 1, *A* is the best strategy if Player 2 plays *C*, and, for Player 2, *C* is the best strategy if Player 1 plays *A*.

Table 1. A payoff matrix for a sample game.

		Player 2	
		C	*D*
Player 1	*A*	(2,1)	(4,0)
	B	(1,0)	(3,2)

Source: J. Pita et al., *Using game theory for Los Angeles Airport security*, "AI Magazine" 2009, vol. 30, no. 1, pp. 43–57.

Players in a non-cooperative game do not have to focus on particular strategies. Instead of choosing a single strategy with certainty, a player can choose one strategy with some probability, other strategy with other probability, and so on and so forth. That is, a player can assign

[9] Pita et al., *Using game theory for Los Angeles Airport...*

probabilities to each available strategy. For instance, Player 1 may choose to play strategy A with a certain probability, denoted by p, and strategy B with a probability of $1 - p$. Similarly, Player 2 can assign probabilities to strategies C and D. By using mixed strategies, the players introduce randomness into their decision-making process. The concept of the Nash equilibrium also extends to the mixed strategies.

Consider the game with the payoff matrix defined in Table 2.

Table 2. An example of payoff matrix for a game with no Nash equilibrium in pure strategies, but with a Nash Equilibrium in mixed strategies.

		Player 2	
		C	D
Player 1	A	(2,1)	(1,2)
	B	(1,2)	(3,1)

In this game, there is no pure strategy Nash equilibrium, since for every strategy profile, it is beneficial for one of the players to switch their strategy, if the other player's strategy is fixed, but there exists a mixed strategy Nash equilibrium, namely the following:

- Player 1's mixed strategy: A with probability $\frac{1}{2}$, B with probability $\frac{1}{2}$;
- Player 2's mixed strategy: C with probability $\frac{2}{3}$, D with probability $\frac{1}{3}$.

The expected payoff for Player 1 in the equilibrium is:

$$\frac{1}{2}\times\frac{2}{3}\times 2+\frac{1}{2}\times\frac{1}{3}\times 1+\frac{1}{2}\times\frac{2}{3}\times 1+\frac{1}{2}\times\frac{1}{3}\times 3=\frac{5}{3}$$

and the expected payoff for Player 2 is:

$$\frac{1}{2}\times\frac{2}{3}\times 1+\frac{1}{2}\times\frac{1}{3}\times 2+\frac{1}{2}\times\frac{2}{3}\times 2+\frac{1}{2}\times\frac{1}{3}\times 1=\frac{3}{2}$$

The above model is, of course, simplified. In particular, in the case of protecting critical infrastructure, one may argue that the players do not move simultaneously. This is because, an attacker may be able to observe the defensive measures (strategies) employed by a defender. To address this problem, let us consider a seminal economic model proposed by Stackelberg[10], a game is played between two players: a leader and a follower.

[10] H. von Stackelberg, *Marktform und Gleichgewicht*, J. Springer 1934.

That is, in contrast to the previous ex-ample, the Stackelberg game is played sequentially rather than simultaneously. Specifically, the leader selects their strategy first, and this choice is observed by the follower, who subsequently determines their own move accordingly.

The Stackelberg model has garnered substantial attention within the realm of security applications due to its inherent ability to capture the dynamics of defender-attacker interactions. In this context, Stackelberg games are often called security games.

The model encompasses the following aspects:

- the defender, who assumes the role of the leader in the Stackelberg game, allocates limited security resources to protect a designated set of targets. Recognizing that adversaries possess the capability to observe defense strategies and exploit discernible patterns, the defender naturally opts for a mixed (randomized) strategy. For instance, in the case of LAX, the management of the canine unit determines the frequency of visits to each terminal by a specific type of patrol within a given week. In other words, they establish the probability distribution for each patrol type across all terminals;
- the attacker, acting as the follower in the Stackelberg game, observes the defender's chosen strategy, i.e., these probability distributions. This assumption embodies a prudent and realistic scenario, presupposing an intelligent attacker who thoroughly surveys critical infrastructure before devising and executing an attack;
- lastly, having acquired knowledge of the probabilities selected by the defender, the attacker strategically selects the optimal course of action for themselves and subsequently executes their move accordingly.

It is crucial to emphasize that the attacker has the capability to observe the probability distribution chosen by the defender but not the defender's actual move. As an illustration, let us consider an scenario involving the United States Coast Guard (USCG) responsible for patrolling the Mexican Bay to combat drug trafficking via boats. The smugglers can observe the frequency of patrols in specific sea areas and how frequently patrol boats alter their course, e.g., by changing the patrol direction to a different one. In other words, the attackers have knowledge of the probability distribution. Nonetheless, they lack the capacity to predict whether a patrol boat will change its course at a given moment or not. So they cannot simply wait until a patrol boat goes away as there is

a non-zero probability that it may immediately return. Interestingly, as we already mentioned in the introduction, the system called PROTECT, based on the Stackelberg game was introduced by the USCG to enhance port/coastal security.

One may say that the defenders of a critical infrastructure are at disadvantage as they move first (decide on the allocation of defense resource and the probability distributions) and their move is observed by the attacker. However, a more careful analysis reveals that the defender, as the first mover, may have a significant influence on the choices made by the attacker. Intuitively, the defender may push the attacker to choose one strategy not the others.

Table 3. A payoff matrix for a sample game.

		Player 2	
		C	*D*
Player 1	*A*	(1,1)	(3,0)
	B	(0,0)	(2,1)

Source: D. Korzhyk et al., *Stackelberg vs. Nash in security games: An extended investigation of interchangeability, equivalence, and uniqueness*, "Journal of Artificial Intelligence Research" 2011, vol. 41, no. 2, pp. 297–327.

As an example, let us assume that Player 1 in Table 3 is the leader in the Stackelberg game. Observe that if the players move simultaneously, then actually the only Nash equilibrium (in pure strategies - observe that, trivially, every Nash equilibrium in pure strategies is also a Nash equilibrium in mixed strategies) of this game is for Player 1 to play A and Player 2 to play C, which gives Player 1 the expected payoff equal to 1. Now, by the power of moving first, Player 1 can choose a uniform mixed strategy of playing A and B with equal probability of 1/2 , instead of A with probability 1 and B with probability 0 as in the case of the Nash equilibrium for a simultaneous game. The choice of the leader pushes the follower (Player 2) to choose strategy D instead of C[11]. In result, by being the leader, Player 1 can secure the expected payoff of 5/2 instead of 1, which is quite a significant difference.

[11] In Stackelberg Games, the assumption is that if the follower remains indifferent, the tie is resolved in favor of the leader, since otherwise the optimal solution is ill-defined.

In Appendix A we present a more formal introduction to security games. Let us now comment on the computational challenges posed by security games and then move on to the overview of approaches in the literature.

Challenges and approaches

The game-theoretic approach described in the previous section has been well-established in the literature for many years. However, it is only in the last two decades that these concepts have been effectively deployed in practice to protect critical infrastructure. The reason for this delayed implementation can be attributed to the computational challenges associated with security games. In this section, we will first discuss these challenges and explore how optimization and AI techniques have emerged as effective tools to address them. Furthermore, we will briefly review the existing lines of research on security games to shed light on the challenges involved in developing practical and feasible solutions.

Computational challenges

In real-life deployments, Stackelberg games pose significant computational challenges due to the following key factors:

- first and foremost, decision spaces in complex and large-scale environments of critical infrastructure are immense. The number of possible strategies and actions that can be taken by the players, such as defenders and attackers, can be truly enormous. For example, the New York City Subway system is one of the largest and busiest public transportation networks in the world, serving millions of commuters and visitors daily. The subway system comprises a vast network of tracks, stations, and interconnected lines, covering a total of 472 stations and over 800 miles (1,287 kilometers) of tracks. There are not only various layers of protective measures (i.e., enormous strategy space of the defender) but also very many attack options (i.e., enormous strategy space of the attacker). This requires efficient algorithms to explore and optimize such a vast decision spaces;
- this difficulty is further exacerbated by uncertainty and incomplete information regarding the intentions, capabilities, and actions

of adversaries. Bayesian security games (see the appendix) explicitly model this uncertainty using probability distributions. This, however, adds computational complexity to the problem;

- next, real-life situations are often dynamic and they constantly evolve. Adversaries may adapt their strategies, and defenders need to respond accordingly. Modeling and optimizing strategies in such dynamic environments require solving repeated or sequential games, which further increase the computational challenge.

The scientific literature took a few routes to deal with the computational challenge posed by security games. One approach is to employ mathematical optimization methods to solve these games efficiently. Researchers have developed algorithms and optimization techniques that can handle large-scale game models and provide solutions within reasonable time frames. These optimization methods exploit the structure of the game to reduce the computational burden and improve computational efficiency. They leverage mathematical programming, linear programming, integer programming, and other optimization frameworks to find optimal strategies and resource allocations.

Due to the inherent complexity of these games, finding exact solutions for large-scale scenarios is often infeasible. Therefore, researchers and practitioners often resort to developing approximation algorithms and heuristics to tackle computational challenges while maintaining a reasonable level of accuracy.

Furthermore, AI techniques may play an important role in enhancing the performance of the optimization algorithms. Using AI to optimize algorithms in general and optimization solvers in particular has led to improvements of the state of the art in the solving of hard computational problems for many years. AI allows existing approaches to scale better, and can be applied in many different contexts, for example the one we consider here, i.e., games and optimization[12]. On the other hand, AI models can approximate the result of expensive computational processes quickly and reliably, allowing to do more with the same amount of resources. For example, when deciding what intervention to use to improve resilience and security, some alternatives will be obviously inferior. AI models can help to identify such inferior interventions quickly and cheaply, even accounting

[12] F. Hutter et al., *Boosting Verification by Automatic Tuning of Decision Procedures*, in: *Proceedings of the 19th International Conference on Computer Aided Verification* (CAV 2007), pp. 27–34.

for the uncertainty of the approximation. The same kind of techniques allow systems like AlphaGO to explore a vast space of possible actions in seconds, used for example to deter wildlife poachers[13].

Another option is to resort to parallel computing and distributed computing techniques. By distributing the computational workload across multiple processors or machines, much larger-scale game models can be handled. Here, advancements in hardware technology that improve parallel computing capabilities, are especially important.

Approaches in the Literature

A brief overview of the Stackelberg games with a couple of illustrative examples can be found in the work by Sinha et al.[14] A major and very recent literature review concerning security games can be found in the paper by Hunt and Zhuang[15]. The review examines the present state-of-the-art in game-theoretic modeling for attacker-defender scenarios and analyzes the literature based on common application areas, modeling approaches, and solution methods, additionally addressing significant gaps in the existing body of research and providing a comprehensive discussion on future directions. Other extensive surveys concerning security gams can be found in the work by Fang and Nguyen[16] and in the one by Nguyen et al.[17], where it is demonstrated that security agencies regularly employ decision aids based on game theory to optimize the allocation of limited security resources against strategic adversaries, and that the unique characteristics of these applications demand innovative solutions from AI systems.

---

[13] S. Gholami et al., *Adversary models account for imperfect crime data: Forecast-ing and planning against real-world poachers*, in: *Proceedings of the 17th International Conference on Autonomous Agents and Multiagent Systems* (AAMAS 2018), pp. 823–831.

[14] A. Sinha et al., *Stackelberg security games: Looking beyond a decade of success*, in: *Proceedings of the Twenty-Seventh International Joint Conference on Artificial Intelligence* (IJCAI 2018), pp. 5494–5501.

[15] K. Hunt, J. Zhuang, *A review of attacker-defender games: Current state and paths forward*, "European Journal of Operational Research" 2023, in press. https://doi.org/10.1016/j.ejor.2023.04.009.

[16] F. Fang, T.H. Nguyen, *Green security games: Apply game theory to addressing green security challenges*, "ACM SIGecom Exchanges" 2016, vol. 15, no. 1, pp. 78–83. https://doi.org/10.1145/2994501.2994507.

[17] T.H. Nguyen et al., *Towards a science of security games*, in: *Mathematical Sciences with Multidisciplinary Applications*, B. Toni (ed.), Springer Cham 2016, pp. 347–381.

Two already classical monographs have emerged as prominent references in the intersection of game theory and security. Tambe's book[18] centers around algorithmic advancements and the adoption of game-theoretic software by government stakeholders. On the other hand, Bier and Azaiez's monograph[19] presents a compilation of works that combine game theory and risk analysis within the realm of security.

The Stackelberg games have been increasingly employed to examine a wide array of security issues, spanning from scenarios like missile defense systems[20], terrorism[21], policing[22], to computer network security[23].

Sometimes the models of security games are being referred to as attacker-defender games. They have been a subject of extensive research for the past years and there is a large body of literature concentrating on variety of different problems. An example, again, might be a resource allocation model where e.g., a government wishes to allocate defensive resources among a set targets (e.g., airports or train stations) in an optimal way, and an adversary seeks to attack some of these targets. In the paper by An et al.[24], the authors present an overview of the aforementioned game-theoretic system PROTECT, utilized by the USCG for scheduling patrols in the Port of Boston and New York (see Image). Importantly, the successful evaluation of PROTECT in the Port of Boston has led to its further deployment in the Port of New York. The foundation of PROTECT lies exactly in the attacker-defender Stackelberg game model. However, the development and implementation of the system involved significant contributions in theory as well as comprehensive evaluations. What

---

[18] M. Tambe, *Security and game theory: algorithms, deployed systems, lessons learned*, Cambridge 2011.

[19] V.M. Bier, M.N. Azaiez, *Game Theoretic Risk Analysis of Security Threats*, Springer 2008. https://doi.org/10.1007/978-0-387-87767-9.

[20] G. Brown et al., *A Two-Sided Optimization for Theater Ballistic Missile Defense*, "Operations Research" 2005, vol. 53, no. 5, pp. 745–763. https://doi.org/10.1287/opre.1050.0231.

[21] T. Sandler, *Terrorism & Game Theory*, "Simulation & Gaming" 2003, vol. 34, no. 3, pp. 319-337. https://doi.org/10.1177/1046878103255492.

[22] N. Gatti et al., *Game theoretical insights in strategic patrolling: Model and algorithm in normal-form*, in: *Proceedings of the 2008 conference on ECAI 2008: 18th European Conference on Artificial Intelligence* (ECAI 2008), pp. 403-407.

[23] K-w. Lye, J. Wing, *Game Strategies in Network Security*, "International Journal of Information Security" 2005, vol. 4, pp. 71–86. https://doi.org/10.1007/s10207-004-0060-x.

[24] B. An et al., *A Deployed Quantal Response-Based Patrol Planning System for the U.S. Coast Guard*, "Interfaces" 2013, vol. 43, no. 5, pp. 400–420. https://doi.org/10.1287/inte.2013.0700.

is crucial about the system, it does not assume that adversaries possess perfect rationality, allowing for more realistic and robust scenarios.

Image. The PROTECT system was deployed by the United States Coast Guards to protect the Staten Island Ferry route operated by the New York City Department of Transportation. The picture shows the United States Coast Guards boat protecting one of the ferry vessels.

It is noteworthy that numerous Stochastic Stackelberg Games exist, wherein the decision-making abilities of the adversary are limited by bounded rationality. The majority of systems based on Stackelberg games have traditionally relied on the conventional game-theoretical assumption of adversaries being perfectly rational, which aligns with the standard in game theory literature. However, this assumption may not accurately reflect the behavior of real-world adversaries, as human adversaries often exhibit bounded rationality. Therefore, taking inspiration from psychological and behavioral economics models, researchers (e.g. Yang et al. [25]) have delved into studying parametrized models of bounded rationality in these games. The models in question offer versatile approaches for incorporating bounded rationality into game settings, making them applicable to a wide range of game interactions beyond Stackelberg Security Games. An example of such an approach is an instance studied in the work by

[25] R. Yang et al., *Improving Resource Allocation Strategy Against Human Adversaries in Security Games*, in: *Proceedings of the Twenty-Second International Joint Conference on Artificial Intelligence* (IJCAI 2011), pp. 458–464.

Nguyen et al.[26], where rather than selecting a single target as the optimal response to the induced coverage C of targets by defense resources, the adversary's response h(C) entails probabilistically choosing a target t based on a probability q_t associated with that target.

Let us conclude by noting that there are various further potential applications of Stackelberg Games in modelling adversarial security scenarios. They do include – among others – the following:

- Patrolling games by Vorobeychik et al.[27], designed to simulate situations where environments need to be patrolled to deter intruders. These games draw inspiration from the widely recognized pursuit-evasion game model but have been expanded in diverse ways, including the incorporation of alarm systems;
- In plan interdiction games by Vorobeychik and Pritchard[28], the defender is tasked with selecting a mitigation strategy to intercept potential attack actions, while the attacker, in response, devises an optimal attack plan that bypasses the implemented mitigations. This model finds relevance in the context of adversaries operating in cybersecurity;
- Audit games (J. Blocki et al.[29]) investigate the economic aspects involved in designing audit mechanisms, with a specific emphasis on efficient resource allocation and suitable punishment schemes. The audit game model expands upon the security game model by introducing an extra parameter related to punishment. These models find practical application in audits aimed at ensuring compliance with privacy policies within diverse institutions, including medical hospitals;

[26] T.H. Nguyen et al., *Analyzing the effectiveness of adversary modeling in security games,* in: *Proceedings of the Twenty-Seventh AAAI Conference on Artificial Intelligence* (AAAI 2013), no. 1, pp. 718–724.

[27] Y. Vorobeychik, B. An, M. Tambe, *Adversarial Patrolling Games,* in*: Papers from the 2012 AAAI Spring Symposium,* vol. 3, pp. 91–98.

[28] Y. Vorobeychik, M. Pritchard, *Plan interdiction games,* in: *Adaptive Autonomous Secure Cyber Systems,* S. Jajodia et al. (ed.), Springer Cham 2020, pp. 159-182. https://doi.org/10.1007/978-3-030-33432-1_8.

[29] J. Blocki et al., *Audit games with multiple defender resources,* in: *Proceedings of the Twenty-Ninth AAAI Conference on Artificial Intelligence* (AAAI 2015), vol. 29, no. 1, pp. 791–797.

- Coalitional security games (Guo et al.[30]) tackle the issue of optimizing the prevention of attacker coalitions, where attackers have the ability to form alliances. This concept is particularly relevant in domains such as disrupting terrorist networks, dismantling cells of these networks, or preventing collusion among multiple attackers.

"AI for Security" Team at the IDEAS NCBR

At the IDEAS NCBR research institute, our team "AI for Security" builds Stackelberg models for various types of critical infrastructure. Currently, we have been focusing on developing software for protecting ports, LNG terminals, railways, and power grids. Figure presents the interface of our basic software.

Figure. The snapshot from the Interface of the security-game software that is under development at IDEAS NCBR by the team "AI for Security". Here, we can see the map of the LNG Terminal in Świnoujście. The red circles represents targets (the size of the circle corresponds to the importance of the target). The blue circles represent the positioning of the patrols and their field of view. The heatmap on the right shows the relative probability of which parts of the site should be patrolled (the optimal strategy of the defender). Note that the positioning of targets and patrols on this snapshot is for demonstration purposes only.

[30] Q. Guo et al., *Coalitional security games*, in: *Proceedings of the 15th International Conference on Autonomous Agents and Multiagent Systems* (AAMAS 2016), pp. 159–167.

Our goal is to create a system with the following features:

- Risk Assessment: our system should conduct continuous risk assessments through the analysis of diverse data sources, current and historical ones, and intelligence reports. In order to optimize the allocation of security resources, it should take into account elements such as threat levels, target desirability, and vulnerabilities.
- Randomized Patrol Strategy: Our system should employ randomized strategies to determine the optimal patrol routes for security personnel coupled with optimal deployment of security devices. By randomizing routes, the system increases the difficulty for potential adversaries to predict security patterns, thereby strengthening the element of surprise and deterring potential threats.
- Dynamic Adaptation: our system should accommodate evolving threats and adapt to changing security scenarios. It should possess the capability to dynamically modify patrol routes and allocate resources in response to real-time information, such as emerging intelligence that updates the knowledge of the actual risk situation. This goal is to ensure optimal coverage and enhance response abilities.
- Collaboration and Coordination: the system should facilitate collaboration among diverse security teams and agencies operating within the protected area. It should enable the sharing of information, coordination of efforts, and real-time intelligence exchange, all aimed at enhancing situational awareness and achieving improved security outcomes.
- Performance Evaluation and Feedback: finally, it should incorporate mechanisms for evaluating performance, enabling security personnel to analyze the effectiveness of the system and adapt strategies accordingly. The system should offers feedback, identifying areas for improvement and recognizing patterns that may warrant attention.

In the recent work, members of our team "AI for Security" published a paper at one of the key computer science conferences: the Conference on Uncertainty in Artificial Intelligence (UAI 2023, Pittsburgh, USA)[31] that analyzed a situation of an attack that has two phases. Typically, the attack

[31] A. Nagórko, P. Ciosmak, T. Michalak, *Two-phase security games*, in*: Proceedings of the Thirty-Nine Conference on Uncertainty in Artificial Intelligence* (UAI 2023), pp. 1489–1498.

in security games is modeled as a one-off assault during which the attacker has no chance to update their strategy even if new valuable information is gained in the process. This, however, does not cover certain tactics that can be applied by ever more agile covert organizations.

To address this, we propose a model in which, in the first phase, the attacker makes a preliminary move designed to gain extra information on the defender's activities in this particular instance of the game. Next, in the second phase, this insight is used to choose an optimal concluding move.

A recent real-world example of the tactics that are explicitly modeled in our two-phase game are the actions of Lukashenko's regime in Belarus which exploits immigrants to probe the border with Ukraine[32]. This callus behaviour puts the lives of the immigrants in extreme danger both due to very difficult terrain and the ongoing war. In more details, Ukraine's northwestern border of nearly 900 km is a heavily forested area full of forbidding wetlands and the Chernobyl Exclusion Zone. On top of that, the border – that was crossed by the Russian army in February 2022 and then subsequently restored by the Ukrainian counteroffensive – is now heavily fortified with trenches, walls and mine fields.

Regrettably, despite the fact that the border has now become one of the most perilous in the world, the Belarusian border guards are actively organizing and coordinating groups of immigrants in an attempt to breach it. Their objective is to expose and disrupt the Ukrainian defenses, which are obligated to respond to such attempts due to the threat posed by Russian saboteurs.

Given that some sophisticated electronic security measures are in place, most of these border crossings are detected. However, it should be noted that detection does not necessarily guarantee the presence of a patrol close enough to pre-vent unauthorized entry, meaning that the border is not entirely impenetrable. Nevertheless, even in cases where a specific section of the border is unguarded at the moment of entry, the Ukrainian headquarters promptly dispatch a team to the area.

Consequently, subsequent attempts to enter the same section of the border are highly unlikely to succeed, given the swift response and reinforcement measures taken by the Ukrainian authorities.

---

[32] V. Romanenko, *Belarus uses migrants for intelligence on the border with Ukraine*, https://www.pravda.com.ua/eng/news/2022/12/6/7379514/ [accessed: 25 VI 2023].

Let us consider a scaled-down version of the problem, with four sections of the Belarus-Ukraine border (S_1, S_2, S_3, and S_4) and two patrol units. This setting can be modelled as a standard security game in the spirit of the one used at the Los Angeles World Airport[33]. Pure strategies (moves) of the Ukrainian defenders are possible assignments of patrols to the sections of the border:

$$I = \{S_1S_2, S_1S_3, S_1S_4, S_2S_3, S_2S_4, S_3S_4\}$$

We assume there are two possible types of the attacker: low- and high-profile human traffickers (type 1 and 2, respectively). The high-profile type of the attacker inflicts a much larger loss upon the defender as they organize much bigger groups. Both types have the same strategy space, i.e., an attacker of each type can either choose one of the four sections of the border or back off, i.e., $J_1 = J_2 = \{S_1, S_2, S_3, S_4, \varnothing\}$. The payoffs of both parties, depending on the attacker type, increase linearly with S_i: for a high-profile attackers payoffs are 50, 100, 150 and 200 respectively and for a low-profile attacker the payoffs are five times smaller. The defender payoffs are opposite, with small random noise added uniformly from interval [–5,5].

Assuming that probabilities of attacks by these two types are $p_1 = 0{,}8$ for the low-profile attacker and $p_2 = 0{,}2$ for the high-profile one, an optimal strategy for the defender is:

$$(x_{S_1S_2}, x_{S_1S_3}, x_{S_1S_4}, x_{S_2S_4}, x_{S_2S_4}, x_{S_3S_4}) = (0\%, 50\%, 0\%, 0\%, 50\%, 0\%)$$

According to this strategy, border sections S_1 and S_2 are never protected simultaneously. Such a situation is typical for Stackelberg equilibria in one-phase games and can be easily exploited by performing a two-phase attack.

Now, let us discuss the concept of a two-phase attack. Assume that, unbeknownst to the defender, the attacker possesses the necessary resources and capabilities of both a low-profile human trafficker and a high-profile one. Consequently, the attacker can attempt to breach two sections of the border sequentially, in distinct phases.

Based on the optimal strategy derived previously, let's consider the scenario where, in the first phase, a low-profile human trafficker makes an attempt to breach the border at section S_1. This initial

[33] J. Pita et al., *Using game theory for Los Angeles Airport...*

phase provides the attacker with valuable information, regardless of the defender's positioning. This is due to the fact that the attacker now possesses knowledge of a conditional probability distribution pertaining to the defender's resources.

Let $t \in \{0\%, 17\%, 33\%, 50\%, 67\%, 83\%, 100\%\}$ be a chance of encountering a two-phase attacker, $(1 - t)$ x 80% be a probability of encountering a low-profile attacker and $(1 - t)$ x 20% be a likelihood of encountering a high-profile attacker. For $t = 0\%$ this is the standard one-phase model, while $t = 100\%$ describes a pure two-phase attack.

Table 4 shows that presence of two-phase attackers significantly alters the Stackelberg equilibrium of the game. For example, for 33% probability of a two-phase attack (with 53% chance of a single-phase low-profile attack and 13% chance of a single-phase high-profile attack, keeping the 4 : 1 low- to high-profile ratio), the optimal defender strategy becomes

$$(x_{S_1S_2}, x_{S_1S_3}, x_{S_1S_4}, x_{S_2S_4}, x_{S_2S_4}, x_{S_3S_4}) = (12\%, 15\%, 17\%, 17\%, 18\%, 21\%)$$

As we see in Table 4, two-phase Stackelberg equilibria are much more robust against changes of attacker profiles.

Table 4. Each row presents an optimal mixed strategy of the defender against a group of attackers with a given chance of encountering a two-phase attack. As we can see in the last row, without presence of two-phase attackers the Stackelberg equilibrium heavily over-fits to the random noise in payoff matrices.

S_1S_2	S_1S_3	S_1S_4	S_2S_3	S_2S_4	S_3S_4	Chance of a two-phase attack
0.085	0.11	0.12	0.2	0.25	0.23	100%
0.085	0.11	0.12	0.2	0.25	0.23	83%
0.12	0.11	0.12	0.2	0.25	0.23	67%
0.12	0.15	0.17	0.17	0.18	0.21	50%
0.12	0.15	0.17	0.17	0.18	0.21	33%
0.15	0.15	0.17	0.16	0.18	0.18	17%
0	0.5	0	0	0.5	0	0%

Moves of the defender (patrol placements)

Table 5 shows how defender payoffs change against different compositions of attacker groups. For example, the expected payoff of the defender against a single-phase attack drops to –175 when single-phase strategy is pitted against a two-phase attacker.

Table 5. Expected defender payoff when playing a strategy from Table 4 against a given chance of a two-phase attack. As we can see in the last column, the loss incurred by playing a strategy that ignores the possibility of a two-phase attack is an order of magnitude larger than over-cautious protection against such attacks.

-16.2	-16.2	-16.2	-20.3	-20.3	-24.9	-175	100%	Chance of a two-phase attack
-14.8	-14.8	-14.8	-17.3	-17.3	-20.9	-146	83%	
-13.4	-13.4	-13.4	-14.3	-14.3	-16.9	-116	67%	
-12	-12	-12	-11.3	-11.3	-12.8	-87.1	50%	
-10.7	-10.7	-10.7	-8.36	-8.36	-8.84	-57.9	33%	
-9.27	-9.27	-9.27	-5.38	-5.38	-4.83	-28.6	17%	
-7.89	-7.89	-7.89	-2.41	-2.41	-0.816	0.7	0%	
100%	83%	67%	50%	33%	17%	0%		

Defender strategy

In order to fix this flaw, we propose a new model which allows for considering one-phase and two-phase attackers simultaneously. With our security model, the expected payoff against coordinated attackers jumps from –175 to –16.2 (the defender is still at a disadvantage). The optimal strategy:

$$(x_{S_1S_2}, x_{S_1S_3}, x_{S_1S_4}, x_{S_2S_4}, x_{S_2S_4}, x_{S_3S_4}) = (8{,}5\%, 11\%, 12\%, 20\%, 25\%, 23\%)$$

forces the low-profile attacker to attack S_1 and the high-profile attacker to back off if S_1 was not patrolled. Note that this comes at a cost: for the uncoordinated (one-phase) attack, when low- and high-profile attackers act independently, this strategy brings payoff –7.89 to the defender (a drop from 0.7).

Summary

In this paper, we have provided an exposition of advanced methodologies to ensure the safety of critical infrastructure that incorporate the combination of game theory, optimization techniques, and AI algorithms. The effectiveness of these methods has been demonstrated through their successful deployment in practice in multiple location in the USA. It is crucial to underline that these improvements were achieved not by adding on security resources but rather by optimally deploying the available ones. The work of our team "AI for Security" at the IDEAS NCBR research institute is focused on extending these results and make them applicable to various types of critical infrastructure and to the security threats that have recently reappeared in Europe. We aspire to have them implemented soon to optimize in practice the protection of the Polish critical infrastructure sites and systems.

A Formal Description

We start with a formal description of security games that follows modern treatment by Xu[34]. Then we describe a broader class of Bayesian Stackelberg games that forms a basis for the two-phase model discussed in the previous part of article and we derive a quadratic optimization problem that can be used to solve these games.

A.1 Security Games

A security game, once again, is a two-player game between a defender and an attacker. The defender possesses multiple security resources and aims to allocate these resources to protect n targets from the set $[n] = \{1,2,\ldots,n\}$. A defender pure strategy is a subset of targets that is protected (covered) in a feasible allocation of these resources. A representation of a pure strategy is a binary vector $e \in \{0,1\}^n$ where the entries of value 1 specify the covered targets. Let $E \subseteq \{0,1\}^n$ denote the set of all defender pure strategies. A defender mixed strategy is a probability distribution x over the elements in E. An attacker pure strategy is a target $i \in [n]$. An attacker mixed strategy is denoted by $y \in \Delta_n$, where Δ_n is an n-dimensional simplex. We will use y_i to denote the probability of attacking target i.

[34] H. Xu, *The Mysteries of Security Games: Equilibrium Computation Be-Comes Combinatorial Algorithm Design*, in: *Proceedings of the 2016 ACM Conference on Economics and Computation* (ACM EC 2016), pp. 497–514.

In a most general phrasing, security games are a form of a bilinear game. A bilinear game is given by a pair of matrices (A,B) and polytopes (P,Q). Given that player 1 plays $x \in P$ and player 2 plays $y \in Q$, the utilities for player 1 and 2 are $x^T Ay$ and $x^T By$ respectively.

We can now give different notions of equilibria for security games. A strategy profile (x, y) is a Nash equilibrium, if:

$$\forall x' \in P \ \forall y' \in Q \ x^T Ay \geq x'^T Ay \ \& \ x^T By \geq x^T By'$$

By The Nash Theorem, there exists at least one NE, possibly multiple NEs, in any bilinear game.

When one player moves before another player, the Stackelberg equilibrium serves as a more appropriate solution concept. A two-player Stackelberg game is played between a leader and a follower. The leader moves first, or equivalently, commits to a mixed strategy. The follower observes the leader's strategy and best responds. The leader's optimal strategy, together with the follower's best response, forms an equilibrium.

Let

$$y_x = \arg\max_{y' \in Q} x^T By'$$

denote the follower's best response to a leader strategy $x \in P$. A strategy profile (x, y) is a strong Stackelberg equilibrium if:

$$x = \arg\max_{x' \in \mathcal{P}} x'^T Ay_{x'} \quad \text{and} \quad y = y_x$$

When $B = -A$, the bilinear game is zero-sum. In such games, both NE and SSE, are equivalent to the minimax equilibrium.

A strategy profile (x, y) is a **minimax equilibrium** if

$$\forall x' \in P \forall y' \in Q \ x^T Ay \geq x'^T Ay \ \& \ x^T Ay \leq x^T Ay'$$

If (x, y) is a minimax equilibrium, the strategy x is the player 1's maximin strategy, and y is the player 2's minimax strategy.

The value of the game is:

$$V = x^T Ay = \max_{x' \in \mathcal{P}} \min_{y' \in Q} x'^T Ay'$$

We will now describe the payoff structure of the game – given that the attacker attacks target i:

- the defender gets a reward r_i if target is covered or a cost c_i if i is uncovered,
- the attacker gets a cost ξ_i if target i is covered or a reward ρ_i if i is uncovered,
- both players have utility 0 on the other $n - 1$ unattacked targets.

A crucial assumption here is the following: for all $i \in [n]$ we have:

$$r_i > c_i \text{ and } \rho_i > \xi_i.$$

This means that:

- covering a target is strictly beneficial to the defender than uncovering it,
- the attacker prefers to attack a target when it is uncovered.

Definition 1 (Security Game). A security game G with n targets is a tuple (r,c,ρ,ξ,E) that satisfies $r_i > c_i$ and $\rho_i > \xi_i$ for all $i \in [n]$.

The defender's utility can be defined as follows:

$$U^d(e,i) = r_i e_i + c_i(1 - e_i)$$

Given $p \in \Delta_{|\mathcal{E}|}$ and $y \in \Delta_n$, the defender's expected utility is:

$$U^d(p,y) = \sum_{e \in \mathcal{E}} \sum_{i \in [n]} p_e\, y_i U^d(e,i) =$$

$$= \sum_{e \in \mathcal{E}} \sum_{i \in [n]} p_e\, y_i \big(r_i e_i + c_i(1 - e_i)\big) =$$

$$= \sum_{i \in [n]} y_i \sum_{e \in \mathcal{E}} p_e \big(r_i e_i + c_i(1 - e_i)\big) =$$

$$= \sum_{i \in [n]} y_i \left(r_i \sum_{e \in \mathcal{E}} p_e\, e_i + c_i \left(1 - \sum_{e \in \mathcal{E}} p_e\, e_i \right) \right)$$

Given $p \in \Delta_{|\mathcal{E}|}$ i and $y \in \Delta_n$ the defender's expected utility is:

$$U^d(p,y) = \sum_{i \in [n]} y_i \left(r_i \sum_{e \in \mathcal{E}} p_e\, e_i + c_i \left(1 - \sum_{e \in \mathcal{E}} p_e\, e_i \right) \right)$$

If we follow the convention of using:

$$x_i := \sum_{e \in \mathcal{E}} p_e \, e_i$$

where x_i is the marginal coverage probability of target i, then we have:

$$U^d(p, y) = \sum_{i \in [n]} y_i \big(r_i x_i + c_i (1 - x_i) \big)$$

Let $x = (x_1, \ldots, x_n)^T$ denote the marginal probability for all targets induced by the mixed strategy p. Then the equation above shows that the defender's expected utility can be compactly expressed as the bilinear form:

$$U^d(x, y) = \sum_{i \in [n]} y_i \big(r_i x_i + c_i (1 - x_i) \big)$$

Note that $U^d(x, y)$ has the bilinear form

$$x^T A y + a x$$

for some non-negative diagonal matrix A.

A note is in order here: the convex hull of E is a polytope of all the feasible (i.e., implementable by a defender mixed strategy) marginal probabilities:

$$\mathcal{P} = \{ x = \sum_{e \in \mathcal{E}} p_e \, e : p \in \Delta_{|\mathcal{E}|} \}$$

so we can simply interpret a point $x \in P$ as a mixed strategy and denote the defender's utility by: $U^d(x, y)$.

Similarly, the attacker's expected utility can be compactly represented in the following form:

$$U^a(x, y) = \sum_{i \in [n]} y_i (\rho_i (1 - x_i) + \xi_i x_i)$$

Note that $U^a(x, y)$ also has the bilinear form

$$x^T B y + \beta y$$

for some non-positive diagonal matrix B.

In zero-sum games, all standard equilibrium concepts are payoff-equivalent to the minimax equilibrium, and our goal is to compute the minimax equilibrium in polynomial time.

When the game is not zero-sum, the main solution concept is the strong Stackelberg equilibrium (SSE): the defender plays the role of the leader and can commit to a mixed strategy before the attacker moves. The attacker observes the defender's mixed strategy and best responds. In this case, the goal is to compute the optimal mixed strategy for the defender to commit to (note that the attacker is not able to observe the defender's real-time deployment, i.e., the sampled pure strategy, since he has to plan the attack before the defender's real-time pure strategy is sampled).

A.2 Bayesian Security Games

The solution deployed at the LAX Airport was based on a broader class of games, called Bayesian security games or Bayesian Stackelberg games. We follow[35] to describe this class and we use the Belarus-Ukraine border protection problem discussed in Section 4 as a running example.

In a Bayesian Stackelberg game, the defender plays against a group of attackers of n distinct types. In each round, the defender plays against a single attacker and encounters the attacker of type $1 \leq t \leq n$ randomly, with probability p_t. Attackers may have different sets of moves at their disposal that inflict different damage to the defender. In the running example, we have a low-profile attacker ($t = 1$) and a high-profile attacker ($t = 2$), with

$$p_1 = 4/5 \text{ and } p_2 = 1/5.$$

We let I denote the set of defender's moves. In the running example, the border patrol assigns two patrolling units to four segments of the border, hence $I = \{S_1S_2, S_1S_3, S_1S_4, S_2S_3, S_2S_4, S_3S_4\}$.

In a Bayesian Stackelberg game, the defender picks his mixed strategy x first. Here $x = \{x_i\}_{(i \in I)}$ is a probability measure on I, which we denote by $x \in Prob(I)$ with

$$Prob(I) = \{x: I \rightarrow R: \sum_{i \in I} x_i = 1 \ , x \geq 0\}$$

Strategy x does not depend on t as the defender doesn't know the type of attacker he will encounter.

[35] A. Nagórko, P. Ciosmak, T. Michalak, *Two-phase security games...*

Let J_t denote the set of moves of attacker of type t. In the running example, $J_1 = J_2 = \{S_1,S_2,S_3,S_4,\varnothing\}$, i.e. attackers may either attack one of the border segments or back off. Attackert picks his strategysecond, with the knowledge of the defender's strategy x.

Although it may seem counter-intuitive at first, it is advantageous to the defender to disclose his mixed strategy to the attacker (but not his current defensive positions). It is quite common to disclose information in such scenarios to force the adversary to a favorable response, see e.g. action "ZNICZ" carried out each year by the Polish Police[36].

In each round of the game, both players move independently, according to strategies x and $y^t(x)$ they picked prior. Let $r_{i,t,j}$ denote the defender's payoff if she played move $i \in I$ against the attacker of type $1 \le t \le n$ who played a move $j \in J_t$. Let $c_{i,t,j}$ denote attacker's payoff (which may be different from $-r_{i,t,j}$) as we do not assume that the games are zero-sum in general.

Player payoffs may be compactly presented using payoff matrices. In the running example, the payoff matrices for the high-profile attack are:

	S_1		S_2		S_3		S_4			$\varnothing$
S_1S_2	51,	-50	102,	-100	-152,	150	-211,	200	0,	0
S_1S_3	55,	-50	-123,	100	175,	-150	-221,	200	0,	0
S_1S_4	59,	-50	-108,	100	-169,	150	206,	-200	0,	0
S_2S_3	-69,	50	101,	-100	168,	-150	-221,	200	0,	0
S_2S_4	-55,	50	113,	-100	-170,	150	212,	-200	0,	0
S_3S_4	-75,	50	-123,	100	166,	-150	211,	-200	0,	0

The first number in row i, column j is the defender payoff $r_{i,t,j}$ (here 1 stands for high-profile attacker $t = 1$). The second number is $c_{i,1,j}$. The payoffs for the low-profile attack are:

		S_1		S_2		S_3		S_4		$\varnothing$
S_1S_2	14,	-10	23,	-20	-34,	30	-42,	40	0,	0
S_1S_3	10,	-10	-20,	20	32,	-30	-43,	40	0,	0
S_1S_4	12,	-10	-23,	20	-33,	30	44,	-40	0,	0
S_2S_3	-11,	10	24,	-20	31,	-30	-41,	40	0,	0
S_2S_4	-11,	10	20,	-20	-31,	30	42,	-40	0,	0
S_3S_4	-11,	10	-21,	20	34,	-30	44,	-40	0,	0

[36] *Policyjne działania Znicz*, https://policja.pl/pol/aktualnosci/210088,Policyjne-dzialania-ZNICZ.html [accessed: 25 VI 2023].

Attacker t picks an optimal strategy $\overline{y}^t = \overline{y}^t(x)$ that depends on strategy x known by him and that maximizes his expected payoff

$$\bar{c} = \sum_{i \in I} \sum_{j \in J_t} x_i \bar{y}_j^t c_{i,t,j}$$

This payoff is maximized by a pure strategy, i.e., y^t is optimal if and only if

$$\overline{c} \geq \sum_{i \in I} x_i c_{i,t,j}$$

The defender acts to maximize his expected payoff against the optimal strategies of the attackers, i.e. he picks an optimal strategy x that maximizes his expected payoff:

$$\sum_{i \in I} \sum_{t=1}^{n} \sum_{j \in J_t} p_t x_i \overline{y}_j^t r_{i,t,j}$$

Hence the following quadratic optimization problem solves Bayesian Stackelberg games:

$$\max_{x, y^t} \sum_{i \in I} \sum_{t=1}^{n} \sum_{j \in J_t} p_t x_i y_j^t r_{i,t,j}$$

subject to:

$$\sum_{i \in I} x_i = 1,$$

$$\sum_{j \in J_t} y_j^t = 1 \text{ for each } 1 \leq t \leq n,$$

$$\sum_{i \in I} \sum_{j \in J_t} x_i y_j^t c_{i,t,j} \geq \sum_{i \in I} x_i c_{i,t,j} \text{ for each } 1 \leq t \leq n, j \in J_t,$$

$$x \geq 0, y^t \geq 0 \text{ for each } 1 \leq t \leq n.$$

This formulation coupled with a linearization technique leads to a mixed integer linear programming formulation of Bayesian Stackelberg games published by Paruchuri et al.[37], as the celebrated DOBSS algorithm.

[37] P. Paruchuri et al., *Playing Games for Security: An Efficient Exact Algorithm for Solving Bayesian Stackelberg Games*, in: *Proceedings of the 7th International Joint Conference on Autonomous Agents and Multiagent Systems* (AAMAS 2008), vol. 2, pp. 895–902.

Bibliography

An B. et al., *A Deployed Quantal Response-Based Patrol Planning System for the U.S. Coast Guard*, "Interfaces" 2013, vol. 43, no. 5, pp. 400–420. https://doi.org/10.1287/inte.2013.0700.

Bier V.M., Azaiez M.N., *Game Theoretic Risk Analysis of Security Threats*, Springer 2008. https://doi.org/10.1007/978-0-387-87767-9.

Blocki J. et al., *Audit games with multiple defender resources*, in: *Proceedings of the Twenty-Ninth AAAI Conference on Artificial Intelligence* (AAAI 2015), vol. 29, no. 1, pp. 791–797.

Brown G. et al., *A Two-Sided Optimization for Theater Ballistic Missile Defense*, "Operations Research" 2005, vol. 53, no. 5, pp. 745–763. https://doi.org/10.1287/opre.1050.0231.

Fang F., Nguyen T.H., *Green security games: Apply game theory to addressing green security challenges*, "ACM SIGecom Exchanges" 2016, vol. 15, no. 1, pp. 78–83. https://doi.org/10.1145/2994501.2994507.

Gatti N. et al., *Game Theoretical Insights in Strategic Patrolling: Model and Algorithm in Normal-Form*, in: *Proceedings of the 2008 conference on ECAI 2008: 18th European Conference on Artificial Intelligence* (ECAI 2008), pp. 403–407. https://doi.org/10.3233/978-1-58603-891-5-403.

Gholami S. et al., *Adversary models account for imperfect crime data: Forecasting and planning against real-world poachers*, in: *Proceedings of the 17th International Conference on Autonomous Agents and Multiagent Systems* (AAMAS 2018), pp. 823–831.

Guo Q. et al., *Coalitional security games*, in: *Proceedings of the 2016 15th International Conference on Autonomous Agents and Multiagent Systems* (AAMAS 2016), pp. 159–167.

Haskell W. et al., *Robust protection of fisheries with COmPASS*, in: *Proceedings of the Twenty-Eighth AAAI Conference on Artificial Intelligence* (AAAI 2014), vol. 28, no. 2, pp. 2978–2983.

Hunt K., Zhuang J., *A review of attacker-defender games: Current state and paths forward*, "European Journal of Operational Research" 2023, in press. https://doi.org/10.1016/j.ejor.2023.04.009.

Hutter F. et al., *Boosting Verification by Automatic Tuning of Decision Procedures*, in: *Proceedings of the 19th International Conference on Computer Aided Verification* (CAV 2007), pp. 27–34.

Korzhyk D. et al., *Stackelberg vs. Nash in security games: An extended investigation of interchangeability, equivalence, and uniqueness*, "Journal of Artificial Intelligence Research" 2011, vol. 41, no. 2, pp. 297–327.

Lye K-w., Wing J., *Game Strategies in Network Security*, "International Journal of Information Security" 2005, vol. 4, pp. 71–86. https://doi.org/10.1007/s10207-004-0060-x.

Nagórko A., Ciosmak P., Michalak T., *Two-phase security games*, in: *Proceedings of the Thirty-Nine Conference on Uncertainty in Artificial Intelligence* (UAI 2023), pp. 1489–1498.

Nguyen T.H. et al., *Analyzing the effectiveness of adversary modeling in security games*, in: *Proceedings of the Twenty-Seventh AAAI Conference on Artificial Intelligence* (AAAI 2013), no. 1, pp. 718–724.

Nguyen T.H. et al., *Towards a science of security games*, in: *Mathematical Sciences with Multidisciplinary Applications*, B. Toni (ed.), Springer Cham 2016, pp. 347–381.

Paruchuri P. et al., *Playing Games for Security: An Efficient Exact Algorithm for Solving Bayesian Stackelberg Games*, in: *Proceedings of the 7th International Joint Conference on Autonomous Agents and Multiagent Systems* (AAMAS 2008), vol. 2, pp. 895–902.

Pita J. et al., *Using game theory for Los Angeles Airport security*, "AI Magazine" 2009, vol. 30, no. 1, pp. 43–57. https://doi.org/10.1609/aimag.v30i1.2173.

Sandler T., *Terrorism & Game Theory*, "Simulation & Gaming" 2003, vol. 34, no. 3, pp. 319337. https://doi.org/10.1177/1046878103255492.

Shieh E. et al., *Protect: A deployed game theoretic system to protect the ports of the United States*, in: *Proceedings of the 11th International Conference on Autonomous Agents and Multi-Agent Systems* (AAMAS 2012), vol. 1, pp. 13–20.

Sinha A. et al., *Stackelberg security games: Looking beyond a decade of success*, in: *Proceedings of the Twenty-Seventh International Joint Conference on Artificial Intelligence* (IJCAI 2018), pp. 5494–5501.

Stackelberg H. von, *Marktform und Gleichgewicht*, J. Springer 1934.

Tambe M., *Security and game theory: algorithms, deployed systems, lessons learned*, Cambridge 2011.

Tsai J. et al., *Iris - a tool for strategic security allocation in transportation networks*, in: *Proceedings of the 8th International Conference on Autonomous Agents and Multi-Agent Systems* (AAMAS 2009, Industry Track), pp. 37–44

Vorobeychik Y., An B., Tambe M., *Adversarial Patrolling Games,* in: *Papers from the 2012 AAAI Spring Symposium,* vol. 3, pp. 91–98.

Vorobeychik Y., Pritchard M., *Plan interdiction games,* in: *Adaptive Autonomous Secure Cyber Systems,* S. Jajodia et al. (ed.), Springer Cham 2020, pp. 159–182. https://doi.org/10.1007/978-3-030-33432-1_8.

Xu H., *The Mysteries of Security Games: Equilibrium Computation Be-Comes Combinatorial Algorithm Design,* in: *Proceedings of the 2016 ACM Conference on Economics and Computation* (ACM EC 2016), pp. 497–514.

Yang R. et al., *Adaptive resource allocation for wildlife protection against illegal poachers,* in: *Proceedings of the 13th International Conference on Autonomous Agents and Multiagent Systems* (AAMAS 2014), pp. 453–460.

Yang R. et al., *Improving Resource Allocation Strategy Against Human Adversaries in Security Games,* in: *Proceedings of the Twenty-Second International Joint Conference on Artificial Intelligence* (IJCAI 2011), pp. 458–464.

Yang R. et al., *Improving resource allocation strategies against human adversaries in security games: An extended study,* "Artificial Intelligence" 2013, vol. 195, pp. 440–469. https://doi.org/10.1016/j.artint.2012.11.004.

Yin Z. et al., *Trusts: Scheduling randomized patrols for fare inspection in transit systems,* in: *Proceedings of the Twenty-Sixth AAAI Conference on Artificial Intelligence* (AAAI 2012), vol. 26, no. 2, pp. 2348–2355.

Zhang Y., Malacaria P., *Bayesian Stackelberg games for cyber-security decision support,* "Decision Support Systems" 2021, vol. 148, art. 113599. https://doi.org/10.1016/j.dss.2021.113599.

Internet sources

Policyjne działania Znicz, https://policja.pl/pol/aktualnosci/210088,Policyjne-dzialania-ZNICZ.html [accessed: 25 VI 2023].

Romanenko V., *Belarus uses migrants for intelligence on the border with Ukraine,* https://www.pravda.com.ua/eng/news/2022/12/6/7379514/ [accessed: 25 VI 2023].

Tomasz P. Michalak, PhD

Leader of an independent research team at IDEAS NCBR and a lecturer at the Faculty of Mathematics, Informatics, and Mechanics at the University of Warsaw. Graduate of the Faculty of Economic Sciences at the University of Warsaw. During his academic career, he conducted research at the Department of Computer Science at the University of Oxford, the School of Electronics and Computer Science at the University of Southampton, the Department of Computer Science at the University of Liverpool, and the Faculty of Applied Economics at the University of Antwerp, where he obtained his Ph.D. in economics.

Michał T. Godziszewski, PhD

Specialist in logic and its applications (in mathematics, philosophy and computer science), artificial intelligence (specialisation - multi-agent systems theory: algorithmic game theory, network analysis, computational social choice theory) and theoretical computer science. His current research focuses on algorithmic analysis of social networks and Stackelberg games, computational complexity in game theory and their applications to security systems modelling.

Andrzej Nagórko, PhD

Adjunct at the Faculty of Mathematics, Informatics and Mechanics at the University of Warsaw. Former employee of the Mathematical Institute of the Polish Academy of Sciences and at universities in United States and Israel. For many years he applied mathematical optimization techniques in diverse fields of mathematics - from artificial intelligence, through game theory to geometric group theory. In IDEAS NCBR he works on applications of these methods to the protection of critical infrastructure.

Game-Theoretic Strategies for Cyber-Physical Infrastructures Under Component Disruptions

Nageswara S. V. Rao, *Fellow, IEEE*, Chris Y. T. Ma, Fei He, *Member, INFORMS*

***Abstract*—Networked infrastructures of recursively defined systems composed of discrete cyber and physical components are considered. The components of basic systems at the finest levels can be disrupted by cyber or physical means, and can be reinforced to survive at certain costs. A problem of ensuring the infrastructure performance is formulated as a game between a provider and an attacker, who probabilistically choose components to reinforce and attack, respectively. The disruptions of this infrastructure are characterized using the aggregate failure correlation function that specifies the conditional failure probability of the infrastructure given the failure of an individual system at that level. The survival probabilities of basic systems satisfy simple product-form, first-order differential equations expressed in terms of the multiplier functions. The utility functions of the provider and attacker are composed of the reward and cost terms, both expressed in terms of the component reinforcement and attack probabilities. The Nash equilibrium of this game is characterized, along with the sensitivity functions of the survival probabilities of basic systems that highlight their dependence individually on the cost-benefit terms, the correlation functions, and the multiplier functions. These results are illustrated using simplified models of a distributed cloud servers infrastructure, a 5G data network infrastructure, a high performance computing federation, and a smart energy grid infrastructure.**

***Index Terms*—recursive system of systems, game theory, networked infrastructure, cloud servers infrastructure, 5G data network, high performance computing federation, energy grid**

I. Introduction

Infrastructures such as distributed cloud server farms, smart energy grids, and high performance computing (HPC) federations are complex networked systems. In addition, their individual systems by themselves are complex with several interdependent cyber and physical parts, and they consist of components at the finest levels. These infrastructures can be modeled as Recursive System of Systems (RSoS) that consist of multiple levels of systems as in [1]. In this paper, we consider their defense against the attacks on strategically chosen components; this formulation is at a finer scale compared to the bulk version in [1], where the attacker and provider choose an integer number of components to attack and reinforce, respectively. A RSoS can be expanded by recursively refining the systems into finer levels to end up at *basic* systems $S_b, b = 1, 2, \ldots, N_{\mathcal{S}}$, that consist entirely of discrete components. For example, in a distributed cloud servers infrastructure of 5 sites each with 10,000 servers, the sites and wide-area network can be modeled as top level systems, and they can be successively refined to reach basic systems with physical components such as computing servers, fiber lines, and network switches and routers, and also cyber components such as virtual machines, containers and operating systems [2].

The components of a basic system S_b must be *operational* by themselves and also be *available* such as being connected to the network to ensure the infrastructure performance. However, the individual components of S_b may be disabled and/or disconnected by cyber and physical disruptions, for example, in the cloud servers infrastructure, by cyber attacks on servers or unintended cuts to fiber lines. The effects of a disruption may propagate to other components within the basic system S_b; for example, cyber attacks may be launched on site servers from a compromised server. In addition, these effects may further propagate to components of other basic systems S_{b_1}, $b_1 \neq b$, due to various correlations. For instance, in the above infrastructure, a physical attack on a fiber line to a wide-area network router may result in all servers at multiple sites being unavailable. Thus, in addition to direct disruptions, the correlations within and between basic systems may render the components unavailable even if they are operational by themselves. In an extreme case, the effects of a component attack may propagate within its basic system and then to others, and eventually degrade the performance of the entire infrastructure.

In response to potential attacks, components may be reinforced to survive the disruptions and continue to provide services; for example, in the above infrastructure, replicated servers and redundant fiber lines ensure the continued service under single component disruptions. However, the component defences incur costs that must be balanced in view of the benefits; for example, in the case of cloud storage defense [3], more frequent checking by the provider reduces the impacts of

N. S. V. Rao is with the Computational Sciences and Engineering Division, Oak Ridge National Laboratory, Oak Ridge, TN 37831, USA. Email: raons@ornl.gov

C. Y. T. Ma is with the Department of Computing, The Hang Seng University of Hong Kong, Hong Kong. Email: chrisma@hsu.edu.hk

F. He is with the Department of Mechanical and Industrial Engineering, Texas A&M University-Kingsville, Kingsville, TX, USA. Email: fei.he@tamuk.edu

Manuscript received April 17, 2022; revised June 30, 2022; accepted August 7, 2022

This manuscript has been co-authored by UT-Battelle, LLC, under contract DE-AC05-00OR22725 with the US Department of Energy (DOE). The US government retains and the publisher, by accepting the article for publication, acknowledges that the US government retains a nonexclusive, paid-up, irrevocable, worldwide license to publish or reproduce the published form of this manuscript, or allow others to do so, for US government purposes. DOE will provide public access to these results of federally sponsored research in accordance with the DOE Public Access Plan (http://energy.gov/downloads/doe-public-access-plan).

0000–0000/00$00.00 © 2022 IEEE

persistent attacks but at an increased cost. These considerations naturally lead to game theoretic formulations [4], [5].

The infrastructure provider is tasked with strategically reinforcing the components against attacks by taking various correlations into account. We consider that components are reinforced and attacked by the provider and attacker of the infrastructure, respectively, using the probabilistic strategies. Let $\mathcal{S}_b = \{c_{b,1}, c_{b,2}, \ldots, c_{b,n_b}\}$ denote the set of n_b discrete components of a basic system S_b. Let $p_{c_{b,k}}$ and $q_{c_{b,k}}$ denote the probability of reinforcement and attack on the component $c_{b,k} \in \mathcal{S}_b$, $k = 1, 2, \ldots, n_b$, respectively. Then, let $P_{S_b} = (p_{c_{b,1}}, p_{c_{b,2}}, \ldots, p_{c_{b,n_b}})^T$ and $Q_{S_b} = (q_{c_{b,1}}, q_{c_{b,2}}, \ldots, q_{c_{b,n_b}})^T$ denote the corresponding reinforcement and attack probability vectors. Let $L_D(P_{S_1}, \ldots, P_{S_{N_\mathcal{S}}})$ and $L_A(Q_{S_1}, \ldots, Q_{S_{N_\mathcal{S}}})$ denote the expected costs of reinforcements and attacks on the basic systems of the entire infrastructure, respectively. We then formulate a game between the provider and attacker wherein the provider minimizes the *utility function*, expressed as the sum of gain and cost product terms [4], given by

$$\begin{aligned} &U_D\left(P_{S_1}, \ldots, P_{S_{N_\mathcal{S}}}, Q_{S_1}, \ldots, Q_{S_{N_\mathcal{S}}}\right) \\ &= F_{D,G}(P_{S_1}, \ldots, P_{S_{N_\mathcal{S}}}, Q_{S_1}, \ldots, Q_{S_{N_\mathcal{S}}}) \\ &\quad \times G_D(P_{S_1}, \ldots, P_{S_{N_\mathcal{S}}}, Q_{S_1}, \ldots, Q_{S_{N_\mathcal{S}}}) \\ &\quad + F_{D,L}(P_{S_1}, \ldots, P_{S_{N_\mathcal{S}}}, Q_{S_1}, \ldots, Q_{S_{N_\mathcal{S}}}) \\ &\quad \times L_D(P_{S_1}, \ldots, P_{S_{N_\mathcal{S}}}), \end{aligned} \tag{1}$$

where $F_{D,G}(\cdot)$ and $F_{D,L}(\cdot)$ are the *reward-multiplier* and *cost-multiplier* functions, respectively, of the provider, and $G_D(\cdot)$ represents the gain or reward of keeping the infrastructure operational. For example, under the advanced persistent threats formulation [6], $G_D(\cdot)$ may represent a quantity such as the data protection level. Similarly, the attacker's *utility function* is given by

$$\begin{aligned} &U_A\left(P_{S_1}, \ldots, P_{S_{N_\mathcal{S}}}, Q_{S_1}, \ldots, Q_{S_{N_\mathcal{S}}}\right) \\ &= F_{A,G}(P_{S_1}, \ldots, P_{S_{N_\mathcal{S}}}, Q_{S_1}, \ldots, Q_{S_{N_\mathcal{S}}}) \\ &\quad \times G_A(P_{S_1}, \ldots, P_{S_{N_\mathcal{S}}}, Q_{S_1}, \ldots, Q_{S_{N_\mathcal{S}}}) \\ &\quad + F_{A,L}(P_{S_1}, \ldots, P_{S_{N_\mathcal{S}}}, Q_{S_1}, \ldots, Q_{S_{N_\mathcal{S}}}) \\ &\quad \times L_A(Q_{S_1}, \ldots, Q_{S_{N_\mathcal{S}}}), \end{aligned}$$

where $F_{A,G}(\cdot)$ and $F_{A,L}(\cdot)$ are the *reward-multiplier* and *cost-multiplier* functions, respectively, of the attacker, and $G_A(\cdot)$ represents the reward of disrupting the operation of the infrastructure.

To develop defense strategies for the provider, we characterize the disruptions at each level of recursion of this infrastructure between a system S_i^l at level l and the rest of the infrastructure that has the same parent system as S_i^l. A system survives when it is both *operational* and *available*. Let N_l be the number of level-l systems, $\mathcal{P}_I(S_i^l)$, or $\mathcal{P}_i^l$ in short, denote the survival probability of a basic or non-basic system S_i^l, $i \in \{1, 2, \ldots, N_l\}$, and $\mathcal{P}_I(\mathcal{S})$ denote the survival probability of the entire RSoS $\mathcal{S}$. For a system S_i^l with its parent system S_j^{l-1}, where $i \in \{1, 2, \ldots, N_l\}$ and $j \in \{1, 2, \ldots, N_{l-1}\}$, the *local aggregate failure correlation* C_i^l is the failure probability of the "rest" of the parent system (namely, S_j^{l-1} without S_i^l) given the failure of S_i^l. Intuitively, it indicates the relative importance of S_i^l by capturing the failure propagation from S_i^l to the other parts of the infrastructure under the same parent; in practice, such correlations are often estimated using the observations or measurements of the infrastructure without requiring a much deeper causal knowledge. For a basic system $S_b = S_i^l$, let $\mathcal{P}_I(S_b)$, or $\mathcal{P}_b$ in short, denote its survival probability, and C_b denote its local aggregate failure correlation. In addition to these system-level correlations, the interdependencies between the components are captured by a first-order differential condition using the *system multiplier function* of a basic system S_b [4]. This two-step characterization is natural in several energy and computing infrastructures [4], and it enables us to focus on critical parts of the infrastructure by "separating" the system- and component-level aspects. The RSoS hierarchy enables us to propagate the correlations across the levels, in particular, from components at different levels to the top infrastructure level. This combination of the *hierarchy and correlations* is a distinguished feature of our formulation, for example, compared to the single-level Stackelberg game of critical infrastructures in [7] and the aggregated cloud storage in [3].

In this paper, we derive the Nash Equilibrium (NE) conditions of this game by minimizing the utility functions with respect to the component attack and reinforcement probabilities [8], and present a succinct characterization of NE. These NE conditions show the dependence of $\mathcal{P}_I(\cdot)$ on cost and benefit terms of the provider, and also the infrastructure correlations and multiplier functions via a single succinct product term $\Delta_{\Pi_b}$for S_b (Lemma 5.1). We also derive a sensitivity function for the survival probability of the basic system $\mathcal{P}_b$ with respect to the component reinforcement probability $p_{c_b,k}$ that highlights the dependence on cost-benefit terms and their partial derivatives, and the correlation and multiplier functions (Theorem 5.1). Interestingly, it also indicates the sensitivity to the expected number of reinforced components of S_b denoted by $\bar{x}_b$, and the results in [1] form a special case.

The critical roles of individual components are better captured in the component-based game formulation of this paper compared to previous formulations based on the number of components [1], [4]; this is so despite the superficial similarities in the expressions for survival probability estimates (details in Section III). At a technical level, this probabilistic game formulation ensures the existence of NE and alleviates the large component size assumption needed for the existence of derivatives in other formulations. The resultant defense strategies are more focused on identifying and accounting for critical components located at possibly different RSoS levels, as illustrated in examples in Section VI.

The organization of this paper is as follows. We briefly describe the related work in Section II. In Section III, we describe a discrete component infrastructure model along with the correlation functions and differential conditions on the system survival probabilities; these definitions are generalizations of previous formulations [1], [4] which are based on the number of components. We present a component-based game formulation in Section IV, and derive NE conditions and sensitivity estimates in Section V. We also describe two special cases, OR systems in Section IV-A and statistically indepen-

dent sub-infrastructures in Section IV-B, wherein the analysis of correlation effects is somewhat simplified. We apply the analytical results to simple models of distributed cloud servers infrastructure, 5G data network infrastructure, HPC federation, and smart energy grid infrastructure in Section VI. We present conclusions in Section VII.

II. Related Work

Critical infrastructures that support energy, communication, computation, and transportation systems are vital to national security and economic well-being. These infrastructures are typically complex networked systems consisting of components that need to be protected against various intentional and incidental disruptions [9]–[12], and are often modeled as complex dynamic systems using partial differential equations to reflect the underlying physical properties [13]. The constituent components in critical infrastructures [11] can be typically categorized as cyber components, such as software for control and communication [2], [14], and physical components, such as hardware of sensors and actuators [15], [16]. Distributed cyber-physical systems (CPSs) typically entail recursive structures, for example, cloud computing infrastructures are composed of distributed servers, which are in turn composed of CPU and GPU nodes; traditional power grids are embedded with microgrids, which are composed of energy generation units, controllable device and energy storage units locally [17]; and autonomous vehicular clouds in which each vehicle is equipped with sensors, computing and data storage resources [18]. The recursive structures of CPSs have been explicitly incorporated into modeling and analyses of system functionality [1].

To counter the potential risks from external attacks and internal incidents in cyber-physical infrastructures [13], [19]–[21], it is critical to develop reinforcement strategies to maintain the functionality of CPS under these disruptions. Reliability and robustness analyses of CPS have been conducted using complex network theory [12], [22], [23], graph-theoretic network dynamics [24]–[26], game-theoretic and probabilistic modeling [6], [9], [27], with application to smart grids [28], computer network infrastructures [6], [29], [30], fuzzy dynamical systems [31], large scale industrial processes [32], cyber-physical control systems [33] and integrated multiple-infrastructure systems [34] (to name a few).

A main goal of an infrastructure provider is to ensure its continuous operation by countering attacks while balancing the defense costs, which naturally leads to game-theoretic utility formulations [35]–[38]. Game theory is typically applied to discover defense strategies by considering adaptive rationality of attackers [27], such as multiple-period games that address multiple time-scales of system dynamics [39]; defense against multiple attack technologies [40]; incomplete information games that account for partial knowledge about the system dynamics and attack [41]–[43]; game-theoretic strategies for resource allocation among multiple targets [44]–[49]; and cumulative prospect theory for subjective attackers and security agents of advanced persistent threats [6], [50]. A comprehensive review of the defense and attack strategies in various system structures using game-theoretic formulations has been presented in [51].

Game theory has also been widely used to improve performance of CPS recently, such as demand-side management [52] in smart grids with integration of electric vehicles and renewable energy [53], defense against cascading failures in power grids [54], intelligent transportation system considering human driving behavior [55], and smart medical services [56].

Although the works above address the functionality of specific infrastructures using game theory, they lack the generality to span multiple levels of complex networked infrastructures. This paper proposes a general framework for the game-theoretic protection of such infrastructures by exploiting the inherent recursive structure and correlations at various levels. Specifically, this paper addresses a class of networked infrastructures with discrete components using utility functions with cost and gain parts. It generalizes and unifies a series of previous works under this formulation in multiple directions: (i) the reinforcements and attacks on individual components generalize the formulations based on the number of components [1], [4], [5], (ii) RSoS characterization generalizes those based on fixed-level systems [57] and components [26], [58], (iii) the utility functions account for component reinforcement and attack probabilities, and encompass both sum and product forms [59], and (iv) the local aggregate correlations extend two-level system correlations to RSoS [4]. Also, this paper expands the illustrative examples of cloud servers infrastructure [60], HPC federation [61], and energy grid infrastructure [4] to encompass this component-based formulation.

From a broader perspective, our approach is related to a number of existing works but differs in certain details, specifically, in terms of objectives, formulation and solution methodology. Our overall performance objective based on incidental disruptions is similar to the reliability and availability characteristics of systems and infrastructures [62], and our analysis is similar to the risk analysis [63] of complex systems; our model, however, is a more generic cyber and physical system compared to the 2-level hierarchical model of multistage interconnect network in [64]. Our formulation of intentional disruptions can be viewed as a generalization of sorts of the byzantine generals problem [65] that considers the message compromises, and our analysis is similar to model-based analysis [66] of complex systems. Our RSoS tree is similar to attack trees [67] and is a specific version of the Markov models and fault trees [63], but is specifically designed to capture the correlations at different levels of the infrastructure. Our solution based on the game theory distinguishes it from these works, and the RSoS model distinguishes it from the single-level Markov models used in game theoretic analyses of electric and smart grids [68], [69].

III. Discrete Component Infrastructure Models

Several critical infrastructures can be modeled as RSoS consisting of multiple levels of systems with discrete components at the finest levels. Examples of distributed cloud servers infrastructure, 5G data network infrastructure, HPC

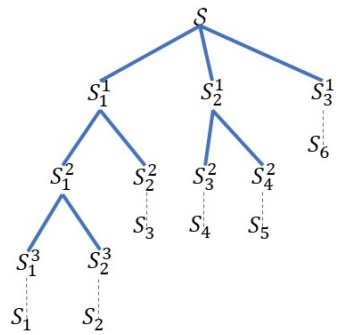


Fig. 1. RSoS tree representation of a simple infrastructure.

federation, and smart energy grid are described in Section VI. An abstract discrete component model has been developed for such infrastructures in terms of correlations and multiplier functions in [4], [5] and extended to RSoS based on the number of components in [1]. We describe extensions of these models in this section to account for the component-based game-theoretic formulation, which is at a finer level than the number of components[1].

A. RSoS Levels and Systems

A RSoS $\mathcal{S}$ [1] is either (i) a *basic* system composed of discrete cyber and physical components, or (ii) composed of N_1 level 1 systems $\mathcal{S}^1 = \{S_i^1 : i = 1, 2, \ldots, N_1\}$ listed from left to right, each of which is a recursively defined system of systems. Thus, $\mathcal{S}$ can be expanded as a *RSoS tree* by refining each non-basic system S_i^1 into next level systems S_j^2's, and so on. By recursively refining, we obtain the basic systems $\mathcal{S}^{\mathrm{b}} = \{S_b : b = 1, 2, \ldots, N_{\mathcal{S}}\}$, located at possibly different finest levels, which correspond to $N_{\mathcal{S}}$ leaves of the RSoS tree located at corresponding depths. We identify the set of systems $\mathcal{S}^l = \{S_i^l : i = 1, 2, \ldots, N_l\}$, listed from left to right at each level $l = 1, 2, \ldots, L$, which may consist of basic and non-basic systems (L is the deepest level of the tree). Thus, for each non-basic system S, we can identify its *descendant* basic systems $\mathcal{B}_S$ as the leaves of the RSoS subtree rooted at S.

A simple RSoS is depicted in Fig. 1 with six basic systems $S_1, S_2, \ldots, S_6$, which are also represented by leaf nodes $S_1^3, S_2^3, S_2^2, S_3^2, S_4^2, S_3^1$, respectively, in the other notation. Notice that RSoS tree structure captures the recursive decomposition of a complex infrastructure into subsystems, and it may not necessarily be a direct reflection of its network topology. In this simple RSoS, we have $\mathcal{S}^2 = \{S_1^2, S_3, S_4, S_5\}$. Here, $\mathcal{B}_{\mathcal{S}} = \mathcal{S}^{\mathrm{b}} = \{S_1, S_2, \cdots, S_6\}$ represents all basic systems of $\mathcal{S}$, and $\mathcal{B}_{S_1^2} = \{S_1, S_2\}$ represents the descendant basic systems of the non-basic system S_1^2.

Our current formulation and that using the number of components [1] are related as follows. For a basic system S_b, the expected number of components reinforced and attacked are denoted by $\bar{x}_b = \sum_{k=1}^{n_b} E\left[1_{D_{c_{b,k}}}\right]$ and $\bar{y}_b = \sum_{k=1}^{n_b} E\left[1_{A_{c_{b,k}}}\right]$, respectively, where $1_{D_{c_{b,k}}}$ and $1_{A_{c_{b,k}}}$ are indicator functions corresponding to the reinforcement and attack of the component $c_{b,k}$, respectively, of the basic system S_b.

[1]For completeness, some of the definitions are presented as in their original form with references, while others are adapted to the current component-based game formulation.

B. Aggregate Correlation Function

At each level l of the infrastructure, we capture the interactions between the systems in terms of survival probabilities using the local aggregate failure correlation function [4]. Let $\mathcal{S}^l_{-S_i^l}$ denote the part of infrastructure at level l without S_i^l that has the same parent system as S_i^l in the RSoS tree, where $i \in \{1, 2, \ldots, N_l\}$, and $\mathcal{P}_I\left(\mathcal{S}^l_{-S_i^l}\right)$ be its survival probability.

Condition 3.1: **Local Aggregate Failure Correlation Function** [4]: The probability that sub-RSoS $\mathcal{S}_j^l$, $j \in \{1, 2, \ldots, N_{l-1}\}$, representing a parent system at level $l-1$ that has level-l system S_i^l, $i \in \{1, 2, \ldots, N_l\}$, as its child system, is operational and available is given by

$$\mathcal{P}_I(\mathcal{S}_j^l) = \mathcal{P}_I(S_i^l) + \mathcal{P}_I\left(\mathcal{S}^l_{-S_i^l}\right) - 1 + C_i^l\left(1 - \mathcal{P}_I(S_i^l)\right), \quad (2)$$

where

$$C_i^l = C_{-i}^l \left(\frac{1 - \mathcal{P}_I\left(\mathcal{S}^l_{-S_i^l}\right)}{1 - \mathcal{P}_I(S_i^l)}\right)$$

is the *local aggregate failure correlation* of system S_i^l. □

Here, $\mathcal{P}_I(\mathcal{S}) = \mathcal{P}_I(\mathcal{S}^1)$ is the survival probability of the infrastructure. This definition can be recursively applied to express $\mathcal{P}_I(S_i^l)$ in terms of $\mathcal{P}_I(\cdot)$'s of descendants of S_i^l and the local aggregate correlations corresponding to level l and below. The local aggregate failure correlation function captures the interdependence of rest of the system $\mathcal{S}^l_{-S_i^l}$ on the failure of S_i^l. In a special case, the failure of S_i^l leads to a definite failure of the rest of the infrastructure; for example, the failure of the communications network renders all supercomputers of a HPC federation unavailable to remote users. Here, we have $C_i^l = 1$ such that $\mathcal{P}_I(\mathcal{S}) = \mathcal{P}_I\left(\mathcal{S}^l_{-S_i^l}\right)$, that is, the survival probability solely depends on $\mathcal{S}^l_{-S_i^l}$.

The effects of component reinforcements on the survival probabilities of systems at various levels are reflected in the partial derivatives with respect to reinforcement probabilities. At each level l, we consider the condition $\frac{\partial \mathcal{P}_I\left(\mathcal{S}^l_{-S_i^l}\right)}{\partial p_{c_{b,k}}} = 0$, where S_b is a descendant basic system of S_i^l with component $c_{b,k}$ (i.e., $S_b \in \mathcal{B}_{S_i^l}$ and $c_{b,k} \in \mathcal{S}_b$). This condition indicates that reinforcement of a component of S_b does not directly impact the survival probability of the rest of RSoS $\mathcal{S}^l_{-S_i^l}$, since S_b is not reachable through the recursive expansion of other systems S_j^l, for $j \neq i$. We capture such system-level considerations at each level for the provider using the following generalization of the condition in [1] to system components.

Condition 3.2: **De-Coupled Reinforcement Aggregate Effects:** For $\mathcal{P}_I(\mathcal{S}_j^l)$, $j = 1, 2, \ldots, N_{l-1}$, in Condition 3.1, we have for systems S_i^l, $i = 1, 2, \ldots, N_l$, of level l

$$\frac{\partial \mathcal{P}_I(\mathcal{S}_j^l)}{\partial p_{c_{b,k}}} = \left(1 - C_i^l\right) \frac{\partial \mathcal{P}_I(S_i^l)}{\partial p_{c_{b,k}}} + \left(1 - \mathcal{P}_I(S_i^l)\right) \frac{\partial C_i^l}{\partial p_{c_{b,k}}}, \quad (3)$$

where S_b is a basic system of S_i^l with component $c_{b,k}$. □

This condition can be recursively applied to express $\frac{\partial \mathcal{P}_I(\mathcal{S})}{\partial p_{c_{b,k}}}$ for the entire infrastructure in terms of $\frac{\partial \mathcal{P}_I(S_b)}{\partial p_{c_{b,k}}}$ and $\mathcal{P}_I(S_b)$ of basic systems, and the aggregate failure correlation functions at levels l through L as will be shown in Lemma 5.1.

C. Basic Systems and Components

A basic system S_b solely consists of components and cannot be further recursively refined. Thus, $\mathcal{P}_I(S_b)$ depends on the correlations between the components within each S_b in addition to its aggregate correlation. These component correlations are characterized by the multiplier function of survival probabilities of basic systems that satisfy the following differential condition.

Condition 3.3: **System Multiplier Functions for Basic Systems:** The survival probability $\mathcal{P}_I(S_b)$ of a basic system S_b, $b = 1, 2, \ldots, N_{\mathcal{S}}$, satisfies the following condition: there exists a *basic system multiplier function* $\Lambda_b(\cdot)$ such that

$$\frac{\partial \mathcal{P}_I(S_b)}{\partial p_{c_{b,k}}} = \Lambda_b \left(P_{X_1}, \ldots, P_{X_{N_{\mathcal{S}}}}, Q_{Y_1}, \ldots, Q_{Y_{N_{\mathcal{S}}}} \right) \mathcal{P}_I(S_b), \tag{4}$$

where $c_{b,k}$ is a component of S_b. □

This condition generalizes an earlier version defined using the number of components [5]. The basic system multiplier functions are defined for basic systems which can be at different levels, whereas Condition 3.2 is defined for systems at the same level l. As an example, consider the contest success functions [70], [71] that model the survival probability of a basic system S_b as $\mathcal{P}_I(S_b) = \frac{\xi + \bar{x}_b}{\xi + \bar{x}_b + \bar{y}_b}$, in terms of the inherent defense level ξ and the expected number of components reinforced and attacked. The basic system multiplier function can then be obtained from the following equation

$$\begin{aligned} \frac{\partial \mathcal{P}_I(S_b)}{\partial p_{c_{b,k}}} &= \frac{\partial \mathcal{P}_I(S_b)}{\partial \bar{x}_b} \frac{\partial \bar{x}_b}{\partial p_{c_{b,k}}} \\ &= \left[\frac{\bar{y}_b}{(\xi + \bar{x}_b + \bar{y}_b)(\xi + \bar{x}_b)} \right] \mathcal{P}_I(S_b), \end{aligned}$$

since $\frac{\partial \bar{x}_b}{\partial p_{c_{b,k}}} = 1$.

We express the survival probability of the infrastructure in term of the correlations and multiplier conditions for all basic systems in the following condition.

Condition 3.4: **Correlations and De-Coupled Reinforcement Conditions for Basic Systems:** For a basic system S_b of $\mathcal{S}$ at level $m_b + 1$ with its parent system $S_{b_{m_b}}^{m_b}$, $b_{m_b} \in \{1, 2, \ldots, N_{m_b}\}$, in the RSoS tree, the conditions corresponding to Conditions 3.1 and 3.2 are given by:

$$\mathcal{P}_I \left(S_{b_{m_b}}^{m_b} \right) = \mathcal{P}_I(S_b) + \mathcal{P}_I \left(\mathcal{S}_{-S_b}^{m_b+1} \right) - 1 + C_b \left[1 - \mathcal{P}_I \left(S_b \right) \right],$$

where C_b is the local aggregate failure correlation function of system S_b, and

$$\frac{\partial \mathcal{P}_I \left(S_{b_{m_b}}^{m_b} \right)}{\partial \bar{x}_b} = (1 - C_b) \frac{\partial \mathcal{P}_I(S_b)}{\partial \bar{x}_b} + (1 - \mathcal{P}_I(S_b)) \frac{\partial C_b}{\partial \bar{x}_b},$$

respectively, for $b = 1, 2, \ldots, N_{\mathcal{S}}$. For the entire infrastructure $\mathcal{S}$, we consider

$$\mathcal{P}_I \left(\mathcal{S} \right) = \mathcal{P}_I(S_b) + \mathcal{P}_I \left(\mathcal{S}_{-S_b} \right) - 1 + \bar{C}_b \left[1 - \mathcal{P}_I \left(S_b \right) \right], \tag{5}$$

where $\bar{C}_b$ is the *global aggregate failure correlation* of S_b and $\mathcal{S}_{-S_b}$ is the rest of the infrastructure without S_b.□

A more general version of this condition can be derived for any part of the infrastructure represented by system S_i^l at level l by expressing $\mathcal{P}_I(S_i^l)$ in terms of the corresponding basic systems $S_b \in \mathcal{B}_{S_i^l}$.

IV. Game Theoretic Formulation

The provider's objective is to sustain infrastructure performance by probabilistically reinforcing the components of basic systems S_b, $b = 1, 2, \ldots, N_{\mathcal{S}}$, as determined by the minimization of the utility function $U_D(\cdot)$. Similarly, the attacker's objective is to disrupt the infrastructure by probabilistically attacking the components of S_b as determined by the minimization of the corresponding utility function $U_A(\cdot)$. For brevity, we only show the derivation of the NE conditions for the provider. However, those for the attacker can be similarly determined.

NE conditions are derived by equating the derivatives of the utility function (Eq. 1) with respect to the component reinforcement probability to zero, which yields [72]

$$\begin{aligned} \frac{\partial U_D}{\partial p_{c_{b,k}}} = \left(G_D \frac{\partial F_{D,G}}{\partial \mathcal{P}_I(\mathcal{S})} + L_D \frac{\partial F_{D,L}}{\partial \mathcal{P}_I(\mathcal{S})} \right) \frac{\partial \mathcal{P}_I(\mathcal{S})}{\partial p_{c_{b,k}}} \\ + F_{D,G} \frac{\partial G_D}{\partial p_{c_{b,k}}} + F_{D,L} \frac{\partial L_D}{\partial p_{c_{b,k}}} = 0 \end{aligned}$$

for $b = 1, 2, \ldots, N_{\mathcal{S}}$, $k = 1, 2, \ldots, n_b$, for the provider. We define the *composite gain-cost* given by

$$L_{G,L}^D \triangleq G_D \frac{\partial F_{D,G}}{\partial \mathcal{P}_I(\mathcal{S})} + L_D \frac{\partial F_{D,L}}{\partial \mathcal{P}_I(\mathcal{S})},$$

which consists of the gain and cost terms scaled with the derivatives of their corresponding multiplier functions with respect to $\mathcal{P}_I(\mathcal{S})$. We also define the *gain-cost gradient* given by

$$\begin{aligned} F_{G,L}^{D,b} &\triangleq F_{D,G} \frac{\partial G_D}{\partial p_{c_{b,k}}} + F_{D,L} \frac{\partial L_D}{\partial p_{c_{b,k}}} \\ &= F_{D,G} \frac{\partial G_D}{\partial \bar{x}_b} + F_{D,L} \frac{\partial L_D}{\partial \bar{x}_b} \end{aligned}$$

which consists of the multiplier functions scaled with the derivatives of their corresponding gain and cost terms with respect to $p_{c_{b,k}}$, $b = 1, 2, \ldots, N_{\mathcal{S}}$, $k = 1, 2, \ldots, n_b$, and to $\bar{x}_b$, as $\frac{\partial \bar{x}_b}{\partial p_{c_{b,k}}} = 1$. Hence, at NE we have the following simplified condition on the partial derivative

$$\frac{\partial \mathcal{P}_I(\mathcal{S})}{\partial p_{c_{b,k}}} = -\frac{F_{G,L}^{D,b}}{L_{G,L}^D} \triangleq -\Theta_b \left(P_{X_1}, \ldots, P_{X_{N_{\mathcal{S}}}}, Q_{Y_1}, \ldots, Q_{Y_{N_{\mathcal{S}}}} \right), \tag{6}$$

where $\Theta_b(\cdot)$ denotes the *scaled gain-cost gradient* of basic system S_b. For brevity, we denote $\Lambda_b \left(P_{X_1}, \ldots, P_{X_{N_{\mathcal{S}}}}, Q_{Y_1}, \ldots, Q_{Y_{N_{\mathcal{S}}}} \right)$ by Λ_b, and $\Theta_b \left(P_{X_1}, \ldots, P_{X_{N_{\mathcal{S}}}}, Q_{Y_1}, \ldots, Q_{Y_{N_{\mathcal{S}}}} \right)$ by Θ_b, for $b = 1, 2, \ldots, N_{\mathcal{S}}$.

We next consider two cases wherein specific forms of correlations lead to simplified NE conditions.

A. Local and Global OR Infrastructures

The OR systems [4] represent infrastructures where simultaneous failures of two or more systems are extremely unlikely at each level. These systems are simpler to analyse due to the absence of system-level correlation terms.

For a *global OR infrastructure*, the failure of a basic system does not impact the rest of the infrastructure $\mathcal{S}$ such that $\bar{C}_b = 0$, for $b = 1, 2, \ldots, N_{\mathcal{S}}$, and hence $\mathcal{P}_I(\mathcal{S}) = \mathcal{P}_I(S_b) + \mathcal{P}_I\left(\mathcal{S}_{-S_b}\right) - 1$. In these (theoretical) systems, for $k = 1, 2, \ldots, n_b$, we have

$$\frac{\partial \mathcal{P}_I(\mathcal{S})}{\partial p_{c_{b,k}}} = \frac{\partial \mathcal{P}_b}{\partial p_{c_{b,k}}} = -\frac{F_{G,L}^{D,b}}{L_{G,L}^{D}} = -\Theta_b.$$

Then, Condition 3.3 (Eq. 4) provides an estimate $\check{\mathcal{P}}_{b;D}$ of the survival probability $\mathcal{P}_I(S_b) = \mathcal{P}_b$ of S_b for the provider, given by

$$\check{\mathcal{P}}_{b;D} = -\frac{\Theta_b}{\Lambda_b},$$

for $b = 1, 2, \ldots, N_{\mathcal{S}}$. This estimate depends mainly on the scaled gain-cost gradient of S_b, and thus represents its "separation" from other basic systems. In this sense, OR systems constitute an important analytical case wherein correlations between the individual systems may be ignored. Although the estimates do not directly involve $\bar{C}_b$, the interactions between basic systems can still be captured by $\Lambda_b(\cdot)$ and $\Theta_b(\cdot)$ which in general depend on other basic systems.

A *local OR infrastructure* is defined by a weaker condition that $C_b = 0$, which implies $\mathcal{P}_I\left(S_{b_{m_b}}^{m_b}\right) = \mathcal{P}_I(S_b) + \mathcal{P}_I\left(\mathcal{S}_{-S_b}^{m_b+1}\right) - 1$.

The cloud memory defense formulation in [3] can be modeled as a single-level OR infrastructure; under the assumption of a single site attack or incidental disruption, the failure of storage at one site does not affect the other ($C_b = 0$), and the sum of the surviving storage can be incorporated into the provider's utility function. We can observe that the survival probability $\check{\mathcal{P}}_{b;D}$ is inversely proportional to the weighted gain term and proportional to its derivative; this seemingly counterintuitive trend is applicable to solutions constrained by Nash equilibria conditions, and not to the unconstrained infrastructure.

B. Statistical Independence of Systems

We consider that the failures of basic system S_b and the rest of the infrastructure $\mathcal{S}_{-S_b}$ are statistically independent such that $\mathcal{P}_I(\mathcal{S}) = \mathcal{P}_I\left(\mathcal{S}_{-S_b}\right)\mathcal{P}_I(S_b)$. Based on Condition 3.4 (Eq. 5) for $\mathcal{S}$, we have $\bar{C}_b = 1 - \mathcal{P}_I\left(\mathcal{S}_{-S_b}\right)$. Using $\frac{\partial \mathcal{P}_I\left(\mathcal{S}_{-S_b}\right)}{\partial p_{c_{b,k}}} = 0$, at NE we have

$$\mathcal{P}_I\left(\mathcal{S}_{-S_b}\right)\frac{\partial \mathcal{P}_I(S_b)}{\partial p_{c_{b,k}}} = -\Theta_b.$$

Applying Condition 3.3 (Eq. 4), at NE we have an estimate for the survival probability of S_b given by

$$\bar{\mathcal{P}}_{b;D} = -\frac{\Theta_b}{\Lambda_b}\frac{1}{\mathcal{P}_I\left(\mathcal{S}_{-S_b}\right)}.$$

Similar to OR systems, the estimate $\bar{\mathcal{P}}_{b;D}$ for statistically independent systems is inversely proportional to the corresponding weighted cost and reward functions, and proportional to their weighted derivatives. However, unlike OR systems, statistical independence is not sufficient to decouple the estimate $\bar{\mathcal{P}}_{b;D}$ from $\mathcal{P}_I(\mathcal{S}_{-S_b})$, i.e., to depend solely on $\Lambda_b(\cdot)$ and $\Theta_b(\cdot)$. In particular, the estimate $\bar{\mathcal{P}}_{b;D}$ is inversely proportional to the survival probability of the rest of the infrastructure at NE.

V. NE Sensitivity Functions: Aggregate Correlations

We now derive estimates for the survival probability $\mathcal{P}_I(S_b)$ of basic system S_b at NE using partial derivatives of the cost and failure correlation functions with respect to $p_{c_{b,k}}$, the reinforcement probability of the component $c_{b,k}$ of S_b. For that purpose, we express the term $\frac{\partial \mathcal{P}_I(\mathcal{S})}{\partial p_{c_{b,k}}}$ by recursively applying Condition 3.2 (Eq. 3), followed by Condition 3.3 (Eq. 4), for the corresponding S_b's, for $b = 1, 2, \ldots, N_{\mathcal{S}}$, $k = 1, 2, \ldots, n_b$.

Lemma 5.1: **Propagation of Correlations and Multiplier Functions:** Consider a component $c_{b,k}$ of the basic system S_b obtained by recursively expanding the RSoS $\mathcal{S}$. Let the path from $\mathcal{S}$ to the leaf S_b at level $m_b + 1$ consists of the sequence of recursively expanding systems $\mathcal{S}, S_{b_1}^1, S_{b_2}^2, ..., S_{b_{m_b}}^{m_b}, S_b$.

(i) By extending the correlation functions and survival probability functions, with Condition 3.2 (Eq. 3) recursively applied until S_b, we have

$$\begin{aligned}\frac{\partial \mathcal{P}_I(\mathcal{S})}{\partial p_{c_{b,k}}} = &\left(1 - C_{\prod_b}\right)\frac{\partial \mathcal{P}_I(S_b)}{\partial p_{c_{b,k}}} \\ &+ \Delta_{\prod_b}\left(\left(1 - \mathcal{P}_I(S_b)\right)\frac{\partial C_b}{\partial p_{c_{b,k}}}\right) + \frac{\partial C_{\Delta_b}}{\partial p_{c_{b,k}}},\end{aligned} \tag{7}$$

where

$$\Delta_{\prod_b} = \prod_{l=1}^{m_b}\left(1 - C_{b_l}^l\right),$$

$$\begin{aligned}\frac{\partial C_{\Delta_b}}{\partial p_{c_{b,k}}} = &\sum_{l=2}^{m_b}\left(\frac{\partial C_{b_l}^l}{\partial p_{c_{b,k}}}\left(1 - \mathcal{P}_I(S_{b_l}^l)\right)\prod_{j=1}^{l-1}\left(1 - C_{b_j}^j\right)\right) \\ &+ \left(1 - \mathcal{P}_I(S_{b_1}^1)\right)\frac{\partial C_{b_1}^1}{\partial p_{c_{b,k}}},\end{aligned}$$

and $C_{\prod_b}$ is the *product aggregate correlation function* given by

$$C_{\prod_b} = 1 - (1 - C_b)\prod_{l=1}^{m_b}\left(1 - C_{b_l}^l\right). \tag{8}$$

(ii) In terms of the *product multiplier function* $\Lambda_{\prod_b}$ of S_b given by

$$\Lambda_{\prod_b} = \frac{\partial \mathcal{P}_I(\mathcal{S})}{\partial \mathcal{P}_I(S_{b_1}^1)}\frac{\partial \mathcal{P}_I(S_{b_1}^1)}{\partial \mathcal{P}_I(S_{b_2}^2)}\cdots\frac{\partial \mathcal{P}_I\left(S_{b_{m_b}}^{m_b}\right)}{\partial \mathcal{P}_I(S_b)}\Lambda_b, \tag{9}$$

we have

$$\frac{\partial \mathcal{P}_I(\mathcal{S})}{\partial p_{c_{b,k}}} = \Lambda_{\prod_b}\mathcal{P}_I(S_b). \tag{10}$$

Proof: To show Part (i), we apply Condition 3.2 (Eq. 3) to $S^1_{b_1}, S^2_{b_2}, ..., S^{m_b}_{b_{m_b}}, S_b$ such that

$$\begin{aligned}
\frac{\partial \mathcal{P}_I(\mathcal{S})}{\partial p_{c_{b,k}}} &= \left(1 - C^1_{b_1}\right) \frac{\partial \mathcal{P}_I(S^1_{b_1})}{\partial p_{c_{b,k}}} + \left(1 - \mathcal{P}_I(S^1_{b_1})\right) \frac{\partial C^1_{b_1}}{\partial p_{c_{b,k}}} \\
\frac{\partial \mathcal{P}_I(S^1_{b_1})}{\partial p_{c_{b,k}}} &= \left(1 - C^2_{b_2}\right) \frac{\partial \mathcal{P}_I(S^2_{b_2})}{\partial p_{c_{b,k}}} + \left(1 - \mathcal{P}_I(S^2_{b_2})\right) \frac{\partial C^2_{b_2}}{\partial p_{c_{b,k}}} \\
\frac{\partial \mathcal{P}_I(S^2_{b_2})}{\partial p_{c_{b,k}}} &= \left(1 - C^3_{b_3}\right) \frac{\partial \mathcal{P}_I(S^3_{b_3})}{\partial p_{c_{b,k}}} + \left(1 - \mathcal{P}_I(S^3_{b_3})\right) \frac{\partial C^3_{b_3}}{\partial p_{c_{b,k}}} \\
\cdots &= \cdots \\
\frac{\partial \mathcal{P}_I\left(S^{m_b}_{b_{m_b}}\right)}{\partial p_{c_{b,k}}} &= (1 - C_b) \frac{\partial \mathcal{P}_I(S_b)}{\partial p_{c_{b,k}}} + (1 - \mathcal{P}_I(S_b)) \frac{\partial C_b}{\partial p_{c_{b,k}}}.
\end{aligned}$$

Then by repeatedly substituting the above equations, we obtain

$$\begin{aligned}
\frac{\partial \mathcal{P}_I(\mathcal{S})}{\partial p_{c_{b,k}}} = &\left(1 - C_{\prod_b}\right) \frac{\partial \mathcal{P}_I(S_b)}{\partial p_{c_{b,k}}} \\
&+ \left(1 - \mathcal{P}_{\prod_b}(S_b)\right) \frac{\partial C_b}{\partial p_{c_{b,k}}} \\
&+ \sum_{l=2}^{m_b} \left(\frac{\partial C^l_{b_l}}{\partial p_{c_{b,k}}} \left(1 - \mathcal{P}_I(S^l_{b_l})\right) \prod_{j=1}^{l-1} \left(1 - C^j_{b_j}\right) \right) \\
&+ \left(1 - \mathcal{P}_I(S^1_{b_1})\right) \frac{\partial C^1_{b_1}}{\partial p_{c_{b,k}}},
\end{aligned}$$

where

$$1 - \mathcal{P}_{\prod_b}(S_b) = (1 - \mathcal{P}_I(S_b)) \prod_{l=1}^{m_b} \left(1 - C^l_{b_l}\right) = \Delta_{\prod_b}(1 - \mathcal{P}_I(S_b)).$$

Part (ii) follows by applying the chain rule to $\frac{\partial \mathcal{P}_I(\mathcal{S})}{\partial p_{c_{b,k}}}$ in terms of $\mathcal{P}_I(\cdot)$'s of $\mathcal{S}, S^1_{b_1}, S^2_{b_2}, ..., S^{m_b}_{b_{m_b}}, S_b$ and then applying Condition 3.3 (Eq. 4). □

Qualitative information about the sensitivities of $\mathcal{P}_I(S_b)$ to different parameters can be inferred using estimates derived in the following result.

Theorem 5.1: **Basic System Survival Probability Estimates:** Under Conditions 3.1, 3.2 and 3.3, the estimates $\hat{\mathcal{P}}_{b;D}$ and $\tilde{\mathcal{P}}_{b;D}$ of the survival probability $\mathcal{P}_I(S_b)$ of basic system S_b expressed in terms of $C_{\prod_b}$ and $\Lambda_{\prod_b}$, respectively, for $b = 1, 2, \ldots, N_{\mathcal{S}}$, are given by

$$\hat{\mathcal{P}}_{b;D} = \frac{\Delta_{\prod_b} \frac{\partial C_b}{\partial p_{c_{b,k}}} + \frac{\partial C_{\Delta_b}}{\partial p_{c_{b,k}}} + \frac{F^{D,b}_{G,L}}{L^D_{G,L}}}{\Delta_{\prod_b} \frac{\partial C_b}{\partial p_{c_{b,k}}} - \left(1 - C_{\prod_b}\right) \Lambda_b} \tag{11}$$

where S_b is a basic system under the condition: $\Delta_{\prod_b} \neq 0$ and $\frac{\partial C_b}{\partial p_{c_{b,k}}} - (1 - C_b)\Lambda_b \neq 0$, and

$$\tilde{\mathcal{P}}_{b;D} = -\frac{1}{\Lambda_{\prod_b}} \frac{F^{D,b}_{G,L}}{L^D_{G,L}}.$$

Proof: Our proof is based on deriving NE conditions for the utility function. At NE (Eq. 6), we have

$$\frac{\partial \mathcal{P}_I(\mathcal{S})}{\partial p_{c_{b,k}}} = -\frac{F^{D,b}_{G,L}}{L^D_{G,L}}.$$

The proof follows along the lines of [1], and we provide the details here for completeness. Using Part (i) of Lemma 5.1 (Eq. 7), we have

$$\begin{aligned}
\frac{\partial \mathcal{P}_I(\mathcal{S})}{\partial p_{c_{b,k}}} &= (1 - C_{\prod_b}) \frac{\partial \mathcal{P}_I(S_b)}{\partial p_{c_{b,k}}} \\
&\quad + \Delta_{\prod_b} \left((1 - \mathcal{P}_I(S_b)) \frac{\partial C_b}{\partial p_{c_{b,k}}} \right) + \frac{\partial C_{\Delta_b}}{\partial p_{c_{b,k}}}, \\
&= -\frac{F^{D,b}_{G,L}}{L^D_{G,L}}.
\end{aligned}$$

Then, by Condition 3.3 (Eq. 4), we have $\frac{\partial \mathcal{P}_I(S_b)}{\partial p_{c_{b,k}}} = \Lambda_b \mathcal{P}_I(S_b)$. Denoting estimate of $\mathcal{P}_I(S_b)$ by $\hat{\mathcal{P}}_{b;D}$, we have

$$\begin{aligned}
(1 - C_{\prod_b})\Lambda_b \hat{\mathcal{P}}_{b;D} + \Delta_{\prod_b} \left(1 - \hat{\mathcal{P}}_{b;D}\right) \frac{\partial C_b}{\partial p_{c_{b,k}}} \\
= -\left(\frac{F^{D,b}_{G,L}}{L^D_{G,L}} + \frac{\partial C_{\Delta_b}}{\partial p_{c_{b,k}}} \right).
\end{aligned}$$

Under the condition $\Delta_{\prod_b} \neq 0$ and $\frac{\partial C_b}{\partial p_{c_{b,k}}} - (1 - C_b)\Lambda_b \neq 0$, we have

$$\Delta_{\prod_b} \frac{\partial C_b}{\partial p_{c_{b,k}}} - \left(1 - C_{\prod_b}\right) \Lambda_b \neq 0,$$

which provides the above expression for $\hat{\mathcal{P}}_{b;D}$. The second estimate $\tilde{\mathcal{P}}_{b;D}$ follows from Part (ii) of Lemma 5.1 (Eq. 10). □

The system survival probability estimate $\hat{\mathcal{P}}_{b;D}$ of basic system S_b provides qualitative information about the effects of various parameters including the product aggregate correlation coefficient $C_{\prod_b}$, basic system multiplier function $\Lambda_b(\cdot)$, composite gain-cost $L^D_{G,L}$ and gain-cost gradient $F^{D,b}_{G,L}$. We note that these estimates may not necessarily lie within range [0,1], and their main purpose here is to illustrate the dependencies of various terms while others are fixed. In particular, $\hat{\mathcal{P}}_{b;D}$ (i) increases and decreases with $F^{D,b}_{G,L}$ and $L^D_{G,L}$, respectively, (ii) increases with $\Lambda_b(\cdot)$, and (iii) depends directly on both C_b and its derivative for $b = 1, 2, \ldots, N_{\mathcal{S}}$, and indirectly on the corresponding correlations and derivatives at different levels of recursion. The effects of $p_{c_{b,k}}$ on the survival probability are captured succinctly by $\Lambda_b(\cdot)$, the failure correlation functions and their derivatives. The component-level correlations of S_b are captured succinctly by the product $\left(1 - C_{\prod_b}\right) \Lambda_b$, with no effect of Λ_b when S_b affects the rest of infrastructure ($C_b = 1$) and full effect for OR systems ($C_b = 0$).

The second estimate $\tilde{\mathcal{P}}_{b;D}$ has a much simpler form but it subsumes correlations at all levels of recursion that correspond to S_b via its product multiplier function $\Lambda_{\prod_b}$, which depends on the partial derivatives of $\mathcal{P}_I(\cdot)$'s at successive levels as indicated in Part (ii) of Lemma 5.1.

By using $\frac{\partial C_b}{\partial p_{c_{b,k}}} = \frac{\partial C_b}{\partial \bar{x}_b}$ and $\frac{\partial C_{\Delta_b}}{\partial p_{c_{b,k}}} = \frac{\partial C_{\Delta_b}}{\partial \bar{x}_b}$ in the expression for $\hat{\mathcal{P}}_{b;D}$, qualitative inferences can be made in terms of $\bar{x}_b$, the expected number of components reinforced. Thus, these survival probability estimates also indicate sensitivity to the expected number of components reinforced.

When applied directly to a RSoS $\mathcal{S}$ that consists of only basic systems S_b (namely, with 0 level of recursion, or $m_b =$

0), this theorem provides a simpler expression derived for SoS than in [72]

$$\hat{\mathcal{P}}_{b;D} = \frac{\frac{\partial C_b}{\partial p_{c_{b,k}}} + \frac{F^{D,b}_{G,L}}{L^D_{G,L}}}{\frac{\partial C_b}{\partial p_{c_{b,k}}} - (1 - C_b)\Lambda_b},$$

wherein we have $\Delta_{\prod_b} = 1$, $\Lambda_{\prod_b} = \Lambda_b$, $C_{\prod_b} = C_b = \bar{C}_b$ and $\frac{\partial C_{\Delta_b}}{\partial p_{c_{b,k}}} = 0$. The added generality of the recursive formulation of RSoS still preserves the basic form of the estimate $\hat{\mathcal{P}}_{b;D}$ valid for the simpler SoS formulation, albeit under the expanded definitions of the product aggregate correlation function and the product multiplier function.

A. Expected Capacity

Let $\mathcal{C}_b \subseteq \mathcal{S}_b$ be a subset of *core* components of the basic system S_b that represent *core* resources of the infrastructure. For example, the *core* resources of distributed cloud servers infrastructure and HPC may represent the capacity of the server computers and computing nodes, respectively, and those of 5G network may represent the number of mobile users or the aggregate data rate. The core resource of the transmission system of a smart energy grid is the capacity of the system, which is determined by considering the topology of the grid. Let $|\mathcal{C}_b|$ be the cardinality of $\mathcal{C}_b$, then the *expected capacity* of the infrastructure is the expected number of surviving core components of the infrastructure given by

$$N_I = \sum_{b=1}^{N_b} |\mathcal{C}_b| \mathcal{P}_b.$$

Based on Theorem 5.1 (Eq. 11), the expected capacity of the infrastructure is defined as [4]

$$\hat{N}_I = \sum_{b=1}^{N_b} \frac{|\mathcal{C}_b| \left(\Delta_{\prod_b} \frac{\partial C_b}{\partial p_{c_{b,k}}} + \frac{\partial C_{\Delta_i}}{\partial p_{c_{b,k}}} + \frac{F^{D,b}_{G,L}}{L^D_{G,L}} \right)}{\Delta_{\prod_b} \frac{\partial C_b}{\partial p_{c_{b,k}}} - \left(1 - C_{\prod_b}\right) \Lambda_b}.$$

For basic systems which do not have the core resources, their effects on the expected capacity of the infrastructure are captured by various forms of correlation functions in the formula.

B. Upper Bound on the Survival Probability

Theorem 5.2: Under Condition 3.1 and Lemma 5.1, an upper bound on the survival probability of the basic system S_b at the Nash equilibrium is

$$\mathcal{P}_I(S_b) < 1 + \frac{\left(L^D_{G,L} \frac{\partial^2 \mathcal{P}_I(\mathcal{S})}{\partial p^2_{c_{b,k}}} + \frac{\partial F^{D,b}_{G,L}}{\partial p_{c_{b,k}}} \right)}{\Delta_{\prod_b} \frac{\partial L^D_{G,L}}{\partial p_{c_{b,k}}} \frac{\partial C_b}{\partial p_{c_{b,k}}}} + \left[(1 - C_{\prod_b}) \frac{\partial \mathcal{P}_I(S_b)}{\partial p_{c_{b,k}}} + \frac{\partial C_{\Delta_b}}{\partial p_{c_{b,k}}} \right] \frac{1}{\Delta_{\prod_b} \frac{\partial C_b}{\partial p_{c_{b,k}}}},$$

where $k = 1, 2, \ldots, n_b$.

Proof: At NE, the first derivative of the utility function (Eq. 1) is given by

$$\frac{\partial U_D}{\partial p_{c_{b,k}}} = \left(G_D \frac{\partial F_{D,G}}{\partial \mathcal{P}_I(\mathcal{S})} + L_D \frac{\partial F_{D,L}}{\partial \mathcal{P}_I(\mathcal{S})} \right) \frac{\partial \mathcal{P}_I(\mathcal{S})}{\partial p_{c_{b,k}}} + F_{D,G} \frac{\partial G_D}{\partial p_{c_{b,k}}} + F_{D,L} \frac{\partial L_D}{\partial p_{c_{b,k}}} = L^D_{G,L} \frac{\partial \mathcal{P}_I(\mathcal{S})}{\partial p_{c_{b,k}}} + F^{D,b}_{G,L},$$

for $b = 1, 2, \ldots, N_{\mathcal{S}}$. The second derivative condition is given by

$$\frac{\partial^2 U_D}{\partial p^2_{c_{b,k}}} = L^D_{G,L} \frac{\partial^2 \mathcal{P}_I(\mathcal{S})}{\partial p^2_{c_{b,k}}} + \frac{\partial L^D_{G,L}}{\partial p_{c_{b,k}}} \frac{\partial \mathcal{P}_I(\mathcal{S})}{\partial p_{c_{b,k}}} + \frac{\partial F^{D,b}_{G,L}}{\partial p_{c_{b,k}}} > 0,$$

which in turn provides a bound on $\frac{\partial \mathcal{P}_I(\mathcal{S})}{\partial p_{c_{b,k}}}$ as follows,

$$\frac{\partial \mathcal{P}_I(\mathcal{S})}{\partial p_{c_{b,k}}} > - \left(L^D_{G,L} \frac{\partial^2 \mathcal{P}_I(\mathcal{S})}{\partial p^2_{c_{b,k}}} + \frac{\partial F^{D,b}_{G,L}}{\partial p_{c_{b,k}}} \right) \Big/ \frac{\partial L^D_{G,L}}{\partial p_{c_{b,k}}}.$$

The upper bound on $\mathcal{P}_I(S_b)$ then follows from Part (i) of Lemma 5.1. □

This theorem indicates that the product aggregate correlation function $C_{\prod_b}$, the partial derivative of the local aggregate failure correlation function $\frac{\partial C_b}{\partial p_{c_{b,k}}}$, and $\Delta_{\prod_b}$ could limit the achievable $\mathcal{P}_I(S_b)$. On the other hand, the achievable $\mathcal{P}_I(S_b)$ at NE can be improved by the cost term $L^D_{G,L}$, and the partial derivatives $\frac{\partial F^{D,b}_{G,L}}{\partial p_{c_{b,k}}}$ and $\frac{\partial C_{\Delta_b}}{\partial p_{c_{b,k}}}$.

VI. Application Examples

We now describe simple models of distributed cloud servers infrastructure, 5G network infrastructure, HPC federation, and smart energy grid infrastructure, and discuss the implications of analytical results derived in the previous section [2].

A. Distributed Cloud Servers Infrastructure

A cloud infrastructure consists of a variety of hardware and software components that are subject to a broad spectrum of attacks and risks, as reflected in the taxonomy in [2]. A distributed cloud servers infrastructure consists of N_F server sites and a user account management system is shown in Fig. 2. Cloud server infrastructures have been modeled using two-level and RSoS models under game formulation using the number of components, instead of the probability of reinforcement and attack on the components, in [60] and [1] respectively. Here, we focus on the component-based game formulation and consider additional software related details in terms of virtual machines and user accounts.

Server sites are connected over a wide-area network which is not explicitly modeled as we assume the wide-area network is not maintained by the provider. At site j, $j = 1, 2, \ldots, N_F$, there are one virtual machine manager, L_j server computers, some routers to connect the computers, some links to connect

[2]Related literature in each of these application areas is extensive, and we only refer to a few examples to contrast them from our hierarchical infrastructure model and the correlations-based game formulation.

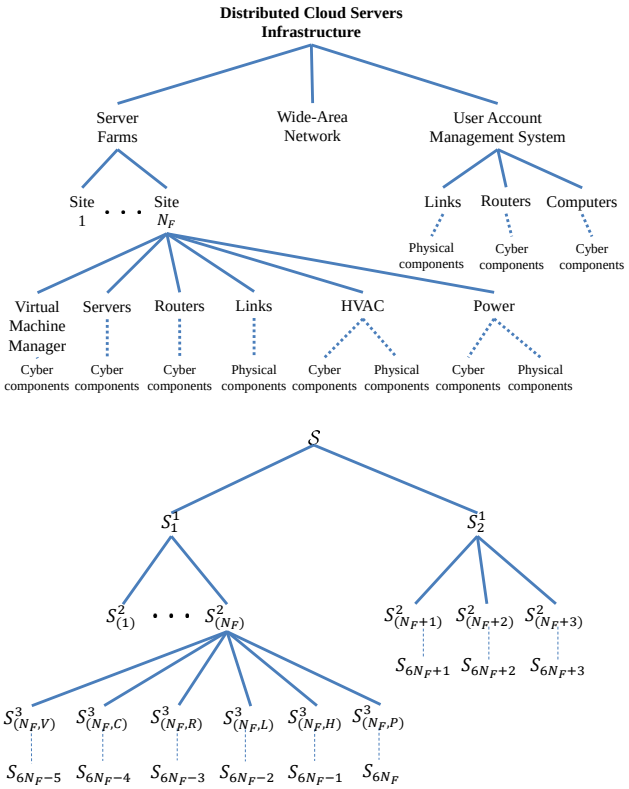


Fig. 2. RSoS tree representation of a distributed cloud servers infrastructure.

the site to the Internet, one Heating, Ventilation and Air-Conditioning (HVAC) system, and one power system. The virtual machine manager, servers, routers, HVAC, and power can be brought down by cyber attacks, and the links, HVAC, and power can be brought down by physical attacks. Meanwhile, the user account management system (UAMS) consists of links to the Internet, routers and also $L_{(U,C)}$ computers. The routers and computers can be brought offline by cyber attacks, while the links can be brought down by physical attacks. To reinforce the components of this infrastructure, redundant components can be deployed, for example, by replicating the servers and routers, and installing redundant fiber lines.

This infrastructure is modeled using $6N_F+3$ basic systems, i.e., $N_{\mathcal{S}} = 6N_F + 3$. At level 2, there are three basic systems, $S^2_{(N_F+1)}$, $S^2_{(N_F+2)}$, and $S^2_{(N_F+3)}$, which represent the links, routers, and computers of the user account management system, respectively. At level 3, there are $6N_F$ basic systems, where $S^3_{(j,V)}$, $S^3_{(j,C)}$, $S^3_{(j,R)}$, $S^3_{(j,L)}$, $S^3_{(j,H)}$, and $S^3_{(j,P)}$ represent the virtual machine manager, server computers, routers, links, HVAC, and power of site j, respectively. Thus, in terms of the original indices, the basic systems can be identified as follows:

(i) *virtual machine manager at a site:* $S_{6j-5} = S^3_{(j,V)}$, for $j = 1, 2, \ldots, N_F$,
(ii) *server computers at a site:* $S_{6j-4} = S^3_{(j,C)}$, for $j = 1, 2, \ldots, N_F$,
(iii) *routers at a site:* $S_{6j-3} = S^3_{(j,R)}$, for $j = 1, 2, \ldots, N_F$,
(iv) *links at a site:* $S_{6j-2} = S^3_{(j,L)}$, for $j = 1, 2, \ldots, N_F$,
(v) *HVAC at a site:* $S_{6j-1} = S^3_{(j,H)}$, for $j = 1, 2, \ldots, N_F$,
(vi) *power at a site:* $S_{6j} = S^3_{(j,P)}$, for $j = 1, 2, \ldots, N_F$,
(vii) *links at UAMS:* $S_{6N_F+1} = S^2_{(N_F+1)}$,
(viii) *routers at UAMS:* $S_{6N_F+2} = S^2_{(N_F+2)}$, and
(ix) *computers at UAMS:* $S_{6N_F+3} = S^2_{(N_F+3)}$.

The relationships between the local aggregate failure correlation functions provide useful insights (as described in [4]), and some of them are listed as follows. For the user account management system, we have

$$C^2_{(N_F+2)} = L_{(U,C)} C^2_{(N_F+1)}$$

which reflects that a successful router attack will disrupt all its $L_{(U,C)}$ connected management computers. (Similarly, a successful link attack will also disrupt its connected management computers.) For the sites, we have $C^3_{(j,L)} = L_j C^3_{(j,C)}$ which indicates that at site j, the fiber disruption will disconnect all its servers. (Similarly, if a virtual machine manager attack, a router attack, a HVAC attack, or a power supply attack is successful, it will also disrupt the operation of the servers at a site.) This multiplicative effect carries over to partial differentials since $\frac{\partial C^3_{(j,L)}}{\partial \mathcal{P}^3_{(j,C)}} = L_j \frac{\partial C^3_{(j,C)}}{\partial \mathcal{P}^3_{(j,C)}}$ and $\frac{\partial C^3_{(j,L)}}{\partial \mathcal{P}^3_{(j,L)}} = L_j \frac{\partial C^3_{(j,C)}}{\partial \mathcal{P}^3_{(j,L)}}$ for $j = 1, 2, \ldots, N_F$, and $\frac{\partial C^2_{(N_F+2)}}{\partial \mathcal{P}^2_{(N_F+3)}} = L_{(U,C)} \frac{\partial C^2_{(N_F+3)}}{\partial \mathcal{P}^2_{(N_F+3)}}$ and $\frac{\partial C^2_{(N_F+2)}}{\partial \mathcal{P}^2_{(N_F+2)}} = L_{(U,C)} \frac{\partial C^2_{(N_F+3)}}{\partial \mathcal{P}^2_{(N_F+2)}}$. Based on Theorem 5.1, this multiplicative effect in partial differentials will be reflected in $\hat{\mathcal{P}}_{(j,L);D}$ and $\hat{\mathcal{P}}_{-(j,L);D}$ in addition to the correlations.

Now consider three basic systems $S_{b_1} = S_{6j-4} = S^3_{(j,C)}$ and $S_{b_2} = S_{6j-2} = S^3_{(j,L)}$ corresponding to the collections of servers and fibers of site j, respectively, and $S_{b_3} = S_{6N_F+1} = S^2_{(N_F+1)}$ corresponding to the links of the user account management system. Then, the product multiplier function (Eq. 9) of S_{b_3} is

$$\Lambda_{\prod_{b_3}} = \frac{\partial \mathcal{P}_I(\mathcal{S})}{\partial \mathcal{P}_I(S^1_2)} \frac{\partial \mathcal{P}_I(S^1_2)}{\partial \mathcal{P}_I\left(S^2_{(N_F+1)}\right)} \Lambda_{b_3}.$$

That of the collection of servers S_{b_1} of site j is

$$\Lambda_{\prod_{b_1}} = \frac{\partial \mathcal{P}_I(\mathcal{S})}{\partial \mathcal{P}_I(S^1_1)} \frac{\partial \mathcal{P}_I(S^1_1)}{\partial \mathcal{P}_I\left(S^2_{(j)}\right)} \frac{\partial \mathcal{P}_I\left(S^2_{(j)}\right)}{\partial \mathcal{P}_I\left(S^3_{(j,C)}\right)} \Lambda_{b_1},$$

which is more complex since S_{b_1} is at a deeper level. Similarly, the product aggregate correlation function (Eq. 8) of S_{b_3} given by

$$C_{\prod_{b_3}} = 1 - (1 - C^1_2)\left(1 - C^2_{(N_F+1)}\right)$$

is simpler than those of S_{b_1} and S_{b_2} given by

$$C_{\prod_{b_1}} = 1 - (1 - C^1_1)\left(1 - C^2_{(j)}\right)(1 - C_{b_1}),$$

$$C_{\prod_{b_2}} = 1 - (1 - C^1_1)\left(1 - C^2_{(j)}\right)(1 - C_{b_2}).$$

Here $C_{b_1} = C_{(j,C)} \approx 0$ since failure of servers does not lead to that of other systems and $C_{b_2} = C_{(j,L)} \approx 1$ since the failure of fibers will render other systems unavailable. Consequently, the product term $\left(1 - C_{\prod_{b_1}}\right)\Lambda_{b_1}$ in Theorem 5.1 indicates higher sensitivity of the servers compared to $\left(1 - C_{\prod_{b_2}}\right)\Lambda_{b_2}$ of the fibers. Note that the two product aggregate correlation functions, $C_{\prod_{b_2}}$ and $C_{\prod_{b_3}}$, are related: (i) C^1_2 relates the

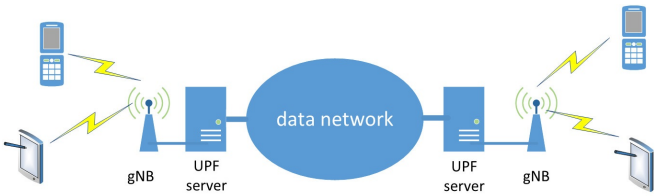


Fig. 3. A 5G network infrastructure.

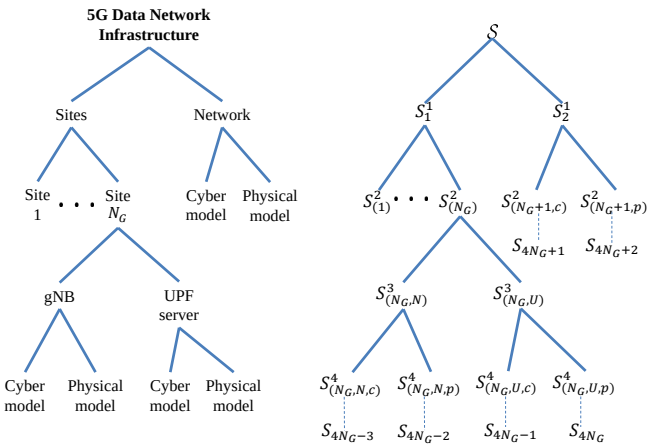


Fig. 4. RSoS tree representation of a 5G network infrastructure.

user account management system S_2^1 to $\mathcal{S}_{-S_2^1}^1$ that subsume $S_{(N_F+1)}^2$ and $S_{(j,L)}^3$, respectively, and (ii) C_1^1 relates the server farm S_1^1 to $\mathcal{S}_{-S_1^1}^1$, which is the user account management system S_2^1, that subsume the site $S_{(j)}^2$ and the link $S_{(N_F+1)}^2$, respectively, and $S_{(j)}^2$ further subsumes $S_{(j,L)}^3$.

The expected capacity of this infrastructure can be specified by the expected number of surviving servers given by $N_I = \sum_{j=1}^{N_F} c_j \mathcal{P}_{6j-4}$, where c_j is the number of servers of the system S_{6j-4} at site j.

B. 5G Network Infrastructure

Recent advances in 5G networking support the provisioning of end-to-end connections between mobile devices with specified Quality of Service (QoS) parameters over a data network using User Plane Functions (UPF) capabilities. This infrastructure and the supported mobile applications are subject to a variety of security threats [73], [74]. In particular, this infrastructure is supported by UPF servers that connect wireless base stations (gNB) to the data network to provide services that complement voice and other services, as shown in Fig. 3, and is also vulnerable to various types of attacks [73]. Here, a gNB serves 5G devices which are within its range, and supports their connections over data network via a UPF server. The RSoS model of a 5G network is shown in Fig. 4. This infrastructure spans N_G sites, $S_1^2, S_2^2, \ldots, S_{N_G}^2$, each with a dedicated gNB and a UPF server. The data network is similar to the network in a cloud servers infrastructure and is represented by its cyber and physical models, denoted by $S_{(N_G+1,c)}^2$ and $S_{(N_G+1,p)}^2$, respectively, and can be further expanded as described in Section VI-A.

The gNB of site i is represented by $S_{(i,N)}^3$ at level 3, and is refined into its cyber model $S_{(i,N,c)}^4$ that captures its cyber attacks, such as infection of malware or malicious firmware, and physical model $S_{(i,N,p)}^4$ that captures the physical attacks, such as physical damage and power disruption, on the cell tower which is typically in an open area. Its UPF server is represented by $S_{(i,U)}^3$ at level 3, and is refined into its cyber model $S_{(i,U,c)}^4$ that captures its cyber attacks, such as malware infection, and physical model $S_{(i,U,p)}^4$ that captures the physical attacks and damages on the fiber connecting it to the gNB. Unlike the gNB which is often in an open area, the UPF server is typically housed inside a facility which is not subject to physical disruption but its fiber connection is still in the open. Reinforcement of the components in such an infrastructure involves deploying redundant servers and fiber connections. Thus, in terms of the original indices, the basic systems can be identified as follows:

(i) *cyber model of wireless base station gNB at a site:* $S_{4j-3} = S_{(j,N,c)}^4$, for $j = 1, 2, \ldots, N_G$,
(ii) *physical model of wireless base station gNB at a site:* $S_{4j-2} = S_{(j,N,p)}^4$, for $j = 1, 2, \ldots, N_G$,
(iii) *cyber model of data network UPF server at a site:* $S_{4j-1} = S_{(j,U,c)}^4$, for $j = 1, 2, \ldots, N_G$,
(iv) *physical model of data network UPF server at a site:* $S_{4j} = S_{(j,U,p)}^4$, for $j = 1, 2, \ldots, N_G$,
(v) *cyber model of the data network:* $S_{4N_G+1} = S_{(N_G+1,c)}^2$, and
(vi) *physical model of the data network:* $S_{4N_G+2} = S_{(N_G+1,p)}^2$.

For the sites, we have $C_{(j,U)}^3 = C_{(j,N)}^3$ which indicates that at site j, the UPF server disruption will disconnect the wireless base station gNB. This effect carries over to partial differentials since $\frac{\partial C_{(j,U)}^3}{\partial \mathcal{P}_{(j,N)}^3} = \frac{\partial C_{(j,N)}^3}{\partial \mathcal{P}_{(j,N)}^3}$ and $\frac{\partial C_{(j,U)}^3}{\partial \mathcal{P}_{(j,U)}^3} = \frac{\partial C_{(j,N)}^3}{\partial \mathcal{P}_{(j,U)}^3}$ for $j = 1, 2, \ldots, N_G$. Based on Theorem 5.1, this effect in partial differentials will be reflected in $\hat{\mathcal{P}}_{(j,U);D}$ and $\hat{\mathcal{P}}_{-(j,U);D}$ in addition to the correlations.

Now consider two basic systems $S_{b_1} = S_{4j-3} = S_{(j,N,c)}^4$ corresponding to the cyber model of gNB at site j and $S_{b_2} = S_{4N_G+2} = S_{(N_G+1,p)}^2$ corresponding to the physical model of the data network. Then, the product multiplier function (Eq. 9) of S_{b_2} is

$$\Lambda_{\prod_{b_2}} = \frac{\partial P_I(\mathcal{S})}{\partial P_I(S_2^1)} \frac{\partial P_I(S_2^1)}{\partial P_I\left(S_{(N_G+1,p)}^2\right)} \Lambda_{b_2}.$$

That of the cyber model of gNB at site j, S_{b_1}, is

$$\Lambda_{\prod_{b_1}} = \frac{\partial P_I(\mathcal{S})}{\partial P_I(S_1^1)} \frac{\partial P_I(S_1^1)}{\partial P_I\left(S_{(j)}^2\right)} \frac{\partial P_I\left(S_{(j)}^2\right)}{\partial P_I\left(S_{(j,N)}^3\right)} \frac{\partial P_I\left(S_{(j,N)}^3\right)}{\partial P_I\left(S_{(j,N,c)}^4\right)} \Lambda_{b_1},$$

which is more complex since S_{b_1} is at a deeper level. Similarly, the product aggregate correlation function (Eq. 8) of S_{b_2} given by

$$C_{\prod_{b_2}} = 1 - \left(1 - C_2^1\right)\left(1 - C_{b_2}\right)$$

is simpler than that of S_{b_1} given by

$$C_{\prod_{b_1}} = 1 - \left(1 - C_1^1\right)\left(1 - C_{(j)}^2\right)\left(1 - C_{(j,N)}^3\right)\left(1 - C_{b_1}\right).$$

Note that these two product aggregate correlation functions are related: (i) C_2^1 relates the network system S_2^1 to $\mathcal{S}_{-S_2^1}^1$

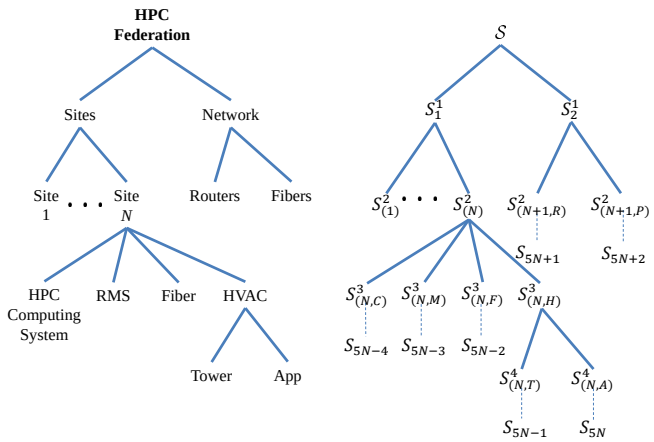


Fig. 5. RSoS tree representation of a HPC federation.

that subsume $S^2_{(N_G+1,p)}$ and $S^4_{(j,N,c)}$, respectively, and (ii) C^1_1 relates the site $S^2_{(j)}$ to $\mathcal{S}^2_{-S^2_{(j)}}$ that subsume $S^4_{(j,N,c)}$ and $S^2_{(N_G+1,p)}$, respectively.

The expected capacity of this 5G data network can be specified in terms of the expected number of surviving gNB given by $N_I = \sum_{j=1}^{N_G} b_j \mathcal{P}_{4j-3}$ where b_j is the capacity of gNB, S_{4j-3}, at site j. Here, the capacity b_j can be defined various ways, including the number of mobile users supported, their aggregate data network capacity and others; in particular, for $b_j = 1$, N_I/N_G represents the fraction of the coverage area served by the surviving gNB.

C. HPC Federation

A generic HPC federation with N geographically distributed sites [75] connected over a wide-area communications network is illustrated in Fig. 5. Each site houses a supercomputer of potentially different size and architecture, a Resource Management System (RMS), and its HVAC system consists of cooling towers and an app that controls its temperature settings. HPC federation can be modeled using RSoS, and its game formulation based on the number of components has been discussed in detail in [61]. In this section, we focus on the component-based formulation with the additional details of RMS.

Cyber attacks on the computing system, RMS, and HVAC control app (typically on a phone or tablet) can bring down the computing system in different ways; for example, the first may crash the computing system, the second may disable the computing system, and the last may manipulate the app to increase the temperature to trigger a system shutdown. Physical damages to the HVAC cooling tower can lead to a computing system shutdown, and cuts of communication fibers can render the computing system unavailable over the networks; these disruptions may be carried out at locations outside the building housing the computing system. Meanwhile, the wide-area communications network connects the sites, and each of its routers manages more than one fiber connection. Let router i manage L_i connections, a successful cyber attack on the router will disconnect all its connections, i.e., the L_i sites, whereas a fiber attack may disconnect one or some of them. To reinforce the components of this infrastructure, some nodes of supercomputers and routers may be replicated, and redundant fiber lines may be installed.

Each site is modeled at a coarse level by five basic systems as shown in Fig. 5, namely a HPC computing system, RMS, site fiber connections, HVAC cooling towers and its app, which can be structured as a sub-tree. As systems, a site and a network are structurally quite different and consist of different types of components, namely, computing and communications devices, respectively. This infrastructure is modeled at a coarse level by a RSoS consisting of $N+2$ systems at level 2, where $S^2_{(N+1,R)}$ and $S^2_{(N+1,P)}$ represent the routers and fiber lines of the communications network as illustrated in Fig. 5. At level 3, there are $4N$ systems where $S^3_{(j,C)}$, $S^3_{(j,M)}$, $S^3_{(j,F)}$, and $S^3_{(j,H)}$ represent the computing system, RMS, fiber connections, and HVAC system of site j, respectively. At level 4, for each site j we have the cooling towers represented by $S^4_{(j,T)}$ and HVAC control app represented by $S^4_{(j,A)}$. Thus, in terms of the original indices, the basic systems at a coarse level can be identified as follows:

(i) *HPC computing systems:* $S_{5j-4} = S^3_{(j,C)}$, for $j = 1, 2, \ldots, N$,
(ii) *RMS:* $S_{5j-3} = S^3_{(j,M)}$, for $j = 1, 2, \ldots, N$,
(iii) *site fiber connections:* $S_{5j-2} = S^3_{(j,F)}$, for $j = 1, 2, \ldots, N$,
(iv) *HVAC system cooling towers and control App:* $S_{5j-1} = S^4_{(j,T)}$, $S_{5j} = S^4_{(j,A)}$ for $j = 1, 2, \ldots, N$, and
(v) *wide-area network with routers and physical connections:* $S_{5N+1} = S^2_{(N+1,R)}$ and $S_{5N+2} = S^2_{(N+1,P)}$.

Thus, there are $N_{\mathcal{S}} = 5N+2$ basic systems that constitute the leaves of the RSoS tree, which correspond to the two basic systems at level 2, $3N$ basic systems at level 3, and $2N$ basic systems at level 4.

For the sites, we have $C^3_{(j,F)} = C^3_{(j,C)}$ which indicates that at site j the fiber disruption will disconnect the HPC computing system. This effect carries over to partial differentials since $\frac{\partial C^3_{(j,F)}}{\partial \mathcal{P}^3_{(j,C)}} = \frac{\partial C^3_{(j,C)}}{\partial \mathcal{P}^3_{(j,C)}}$ and $\frac{\partial C^3_{(j,F)}}{\partial \mathcal{P}^3_{(j,F)}} = \frac{\partial C^3_{(j,C)}}{\partial \mathcal{P}^3_{(j,F)}}$ for $j = 1, 2, \ldots, N$. Based on Theorem 5.1, this effect in partial differentials will be reflected in $\hat{\mathcal{P}}_{(j,F);D}$ and $\hat{\mathcal{P}}_{-(j,F);D}$ in addition to the correlations. Similarly, we have $C^3_{(j,M)} = C^3_{(j,C)}$ which indicates that at site j, the RMS disruption will also disconnect the HPC computing system, and this effect also carries over to partial differentials.

Now consider two basic systems $S_{b_1} = S_{4j} = S^4_{(j,A)}$ corresponding to the HVAC app of site j and $S_{b_2} = S_{4N+2} = S^2_{(N+1,P)}$ corresponding to the fiber lines of the communications network. Then, the product multiplier function (Eq. 9) of S_{b_2} is

$$\Lambda_{\prod_{b_2}} = \frac{\partial P_I(\mathcal{S})}{\partial P_I(S^1_2)} \frac{\partial P_I(S^1_2)}{\partial P_I\left(S^2_{(N+1,P)}\right)} \Lambda_{b_2}.$$

That of the HVAC app of site j, S_{b_1}, is

$$\Lambda_{\prod_{b_1}} = \frac{\partial P_I(\mathcal{S})}{\partial P_I(S^1_1)} \frac{\partial P_I(S^1_1)}{\partial P_I\left(S^2_{(j)}\right)} \frac{\partial P_I\left(S^2_{(j)}\right)}{\partial P_I\left(S^3_{(j,H)}\right)} \frac{\partial P_I\left(S^3_{(j,H)}\right)}{\partial P_I\left(S^4_{(j,A)}\right)} \Lambda_{b_1},$$

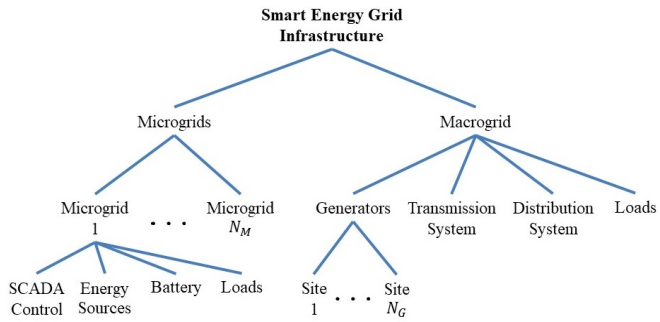


Fig. 6. Representation of a smart energy grid infrastructure.

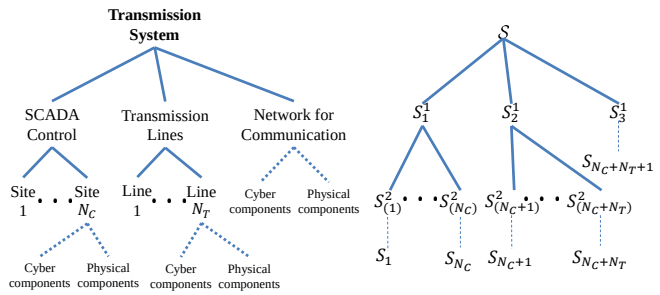


Fig. 7. RSoS tree representation of a transmission system in the macrogrid.

which is more complex since S_{b_1} is at a deeper level. Similarly, the product aggregate correlation function (Eq. 8) of S_{b_2} given by

$$C_{\prod_{b_2}} = 1 - \left(1 - C_2^1\right)\left(1 - C_{b_2}\right)$$

is simpler than that of HVAC app of site j, S_{b_1}, given by

$$C_{\prod_{b_1}} = 1 - \left(1 - C_1^1\right)\left(1 - C_{(j)}^2\right)\left(1 - C_{(j,H)}^3\right)\left(1 - C_{b_1}\right).$$

Note that these two product aggregate correlation functions are related: (i) C_2^1 relates the network system S_2^1 to $\mathcal{S}_{-S_2^1}^1$ that subsume $S_{(N+1,P)}^2$ and $S_{(j,A)}^4$, respectively, and (ii) C_1^1 relates the site $S_{(j)}^2$ to $\mathcal{S}_{-S_{(j)}^2}^2$ that subsume $S_{(j,A)}^4$ and $S_{(N+1,P)}^2$, respectively. Here, $C_{b_1} \approx 1$ since the failure of HVAC app at site j renders its computing system unavailable, and $C_{b_2} \approx 1$ since failures of all fibers in the wide-area network render all computing systems unavailable, which in turn lead to $C_{\prod_{b_2}} \approx 1$; consequently both $C_{\prod_{b_1}}$ and $C_{\prod_{b_2}}$ approach 1.

Let c_j and g_j represent the number of CPU and GPU processor cores, respectively, at the computing system S_{4j-3} at site j. The expected capacity of this HPC federation can be specified by the expected number of surviving processor cores given by $N_I = \sum_{j=1}^{N}(c_j + g_j)\mathcal{P}_{4j-3}$.

D. Smart Energy Grid Infrastructure

A smart grid infrastructure may consist of a macrogrid that provides stable long-term energy sources, which complements a collection of smaller-scale, dynamic microgrids [17], [76]. Typically, a microgrid consists of renewable power generators such as solar panels and wind turbines, smart meters and energy storage devices such as batteries, to support small systems such as homes and offices. A game theoretic formulation of a microgrid infrastructure has been studied using the prospect theory in [77], which can be viewed to constitute a single-level RSoS model. We consider a simplified model of a smart energy grid infrastructure as shown in Fig. 6, which consists of one macrogrid and N_M microgrids. It is an expanded version of a much simpler two-level model of a smart power grid studied in [5]. For simplicity, we focus only on the transmission system of the macrogrid as shown in Fig. 7, which is a main part of the smart energy grid infrastructure; the microgrids can be similarly analyzed. The transmission system is controlled by N_C Supervisory Control and Data Acquisition (SCADA) sites, such that site j controls the power flow on L_j transmission lines, for $j = 1, 2, \ldots, N_C$. The SCADA system of site j may become non-operational due to a direct cyber or physical attack, which in turn will disrupt power flow on all its L_j lines. The SCADA systems are connected to a communications network. Hence, the SCADA system of site j may also become non-operational due to a direct cyber or physical attack of the communications network. A transmission line may also become non-operational due to a direct cyber or physical attack.

This transmission system is modeled using $N_C + N_T + 1$ basic systems, i.e., $N_{\mathcal{S}} = N_C + N_T + 1$. At level 1, there is one basic system, S_3^1, which represents the communications network. At level 2, there are $N_C + N_T$ basic systems, where $S_{(j)}^2$, $j = 1, 2, ..., N_C$, represent the model of SCADA sites and $S_{(N_C+j)}^2$, $j = 1, 2, ..., N_T$, represent the transmission lines. Thus, in terms of the original indices, the basic systems can be identified as follows:

(i) *SCADA site model:* $S_j = S_{(j)}^2$, for $j = 1, 2, \ldots, N_C$,
(ii) *transmission line model:* $S_{N_C+j} = S_{(N_C+j)}^2$, for $j = 1, 2, \ldots, N_T$, and
(iii) *network model:* $S_{N_C+N_T+1} = S_3^1$.

For this transmission system, we have $C_{(j)}^2 = L_j C_{(N_C+m)}^2$, for $j = 1, 2, \ldots, N_C$ and $m = 1, 2, \ldots, N_T$, since a successful cyber attack on the SCADA system at site j would disrupt the power flow on all L_j lines. Similarly, a successful communications network attack would also disrupt the power flow on all transmission lines by disrupting the communication between the transmission lines and SCADA systems, i.e., we have $C_3^1 = N_C C_{(j)}^2 = N_C \times L_j C_{(m)}^2 = N_T C_{(N_C+m)}^2$, for $j = 1, 2, \ldots, N_C$ and $m = 1, 2, \ldots, N_T$. Also, these multiplicative effects will be reflected in the corresponding partial derivatives, and in $\hat{\mathcal{P}}_{i;D}$ and $\hat{\mathcal{P}}_{-i;D}$, for $i = 1, 2, \ldots, (N_C + N_T + 1)$.

Now consider two basic systems $S_{b_1} = S_3^1$ corresponding to the communications network and $S_{b_2} = S_{(N_C+1)}^2$ corresponding to a transmission line. Then, the product multiplier function (Eq. 9) of the communications network S_{b_1} is

$$\Lambda_{\prod_{b_1}} = \frac{\partial \mathcal{P}_I(\mathcal{S})}{\partial \mathcal{P}_I(S_3^1)} \Lambda_{b_1},$$

while that of the transmission line S_{b_2} is

$$\Lambda_{\prod_{b_2}} = \frac{\partial \mathcal{P}_I(\mathcal{S})}{\partial \mathcal{P}_I(S_2^1)} \frac{\partial \mathcal{P}_I(S_2^1)}{\partial \mathcal{P}_I\left(S_{(N_C+1)}^2\right)} \Lambda_{b_2},$$

which is more complex since S_{b_2} is at a deeper level. Similarly, the product aggregate correlation function (Eq. 8) of S_{b_1} given by

$$C_{\prod_{b_1}} = 1 - (1 - C_3^1) = C_{b_1}$$

is simpler than that of S_{b_2} given by

$$C_{\prod_{b_2}} = 1 - (1 - C_2^1)(1 - C_{b_2}).$$

Note that these two product aggregate correlation functions are related: (i) $C_{b_1} = C_3^1$ relates the communications network S_3^1 to $\mathcal{S}^1_{-S_3^1}$ which subsumes $S^2_{(N_C+1)}$, and (ii) C_2^1 relates the transmission line $S^2_{(N_C+1)}$ to $\mathcal{S}^2_{-S^2_{(N_C+1)}}$ which subsumes S_3^1.

Let f_j represent the capacity of flow in transmission line S_j, then the expected total flow in this infrastructure is given by $N_I = \sum_{j=1}^{N_T} f_j \mathcal{P}_{N_C+j}$. With suitable Boolean values assigned to f_j's, the number of the surviving microgrids and SCADA sites can be obtained.

VII. Conclusions

A class of infrastructures with recursively defined systems with discrete cyber and physical components are studied, under a game theoretic formulation with sum-of-products utility functions. Individual components of a system can be disrupted directly or indirectly by either by a cyber or physical attack. The components may be reinforced against such attacks by explicitly taking into account the correlations between the systems and also between the components within individual systems. The interactions between the systems in these infrastructures are characterized using the aggregate correlation functions, and those between the components within a system are characterized by the system multiplier functions; both functions can be estimated from observations or measurements, or analytically modeled in practical scenarios. By formulating a game between an infrastructure provider and an attacker, we derived the Nash Equilibrium conditions in terms of the partial derivatives of cost terms, aggregate correlation functions and survival probabilities of component systems and their partial derivatives. We then estimated the sensitivity functions that indicate the dependence of the system survival probability on these quantities.

These results address the attacks and reinforcements of individual components, and thus extend the previous results that model interactions using the number of components attacked and reinforced in each basic system [1], [4]. In particular, these results enable us to derive cost-centric strategies using simple models of distributed cloud servers, 5G data networks, high performance computing federations, and smart energy grid infrastructures.

Several extensions of the formulation studied in this paper can be pursued in future studies. It would be interesting to study the sequential game formulations of this problem [78], and cases where different levels and types of knowledge are available to attackers and providers, including, misinformation, persistent threats and subjective attacker priorities. Another research direction is to include explicit models for the survival of systems [79] and their interdependencies [57], and subjective decision making using the prospect theory [3], [6]. Applications of our approach to more detailed models of distributed cloud servers infrastructure, 5G data network infrastructures, high performance computing federations, smart energy grid infrastructures, and other infrastructures (such as metro systems [5]) would be of future interest. In several infrastructures, the wide-area networks play a dual, often conflicting, role in providing the critical accessibility to services and also as potential conduits for cyber attacks; it would be of future interest to develop formulations that explicitly reflect these aspects. It would also be of future interest to explore the applicability of this overall method to infrastructures characterized by continuous models such as partial differential equations describing the individual systems or the entire infrastructure, for example, transportation systems [80].

Acknowledgments

The authors acknowledge the constructive feedback from the anonymous reviewers that significantly improved the presentation and references to related works in this paper. This work is funded by RAMSES and Cyber-Physical Networks projects, Office of Advanced Computing Research, U.S. Department of Energy, and by Extreme Scale Systems Center, sponsored by U.S. Department of Defense, and performed at Oak Ridge National Laboratory managed by UT-Battelle, LLC for U.S. Department of Energy under Contract No. DE-AC05-00OR22725; and partially supported by grant UGC/FDS14/E01/19 from the Research Grants Council of the Hong Kong Special Administrative Region and the Deep Learning and Cognitive Computing Centre in The Hang Seng University of Hong Kong.

References

[1] N. S. V. Rao, C. Y. T. Ma, and F. He, "On defense strategies for recursive system of systems using aggregated correlations," in *International Conference on Information Fusion*, 2018.

[2] N. V. Juliadotter and K.-K. R. Choo, "Cloud attack and risk assessment taxonomy," *IEEE Cloud Computing*, vol. 2, no. 1, pp. 14–20, 2015.

[3] L. Xiao, D. Xu, C. Xie, N. B. Mandayam, and H. V. Poor, "Cloud storage defense against advanced persistent threats: A prospect theoretic study," *IEEE Journal on Selected Areas in Communications*, vol. 35, no. 3, pp. 534–544, 2017.

[4] N. S. V. Rao, C. Y. T. Ma, K. Hausken, F. He, D. K. Y. Yau, and J. Zhuang, "Defense strategies for asymmetric networked systems with discrete components," *Sensors*, vol. 18, p. 1421, 2018.

[5] N. S. V. Rao, C. Y. T. Ma, F. He, D. K. Y. Yau, and J. Zhuang, "Cyberphysical correlation effects in defense games for large discrete infrastructures," *Games*, vol. 9, no. 52, 2018.

[6] L. Xiao, D. Xu, N. B. Mandayam, and H. V. Poor, "Attacker-centric view of a detection game against advanced persistent threats," *IEEE transactions on mobile computing*, vol. 17, no. 11, pp. 2512–2523, 2018.

[7] Y. Li, S. Qiao, Y. Deng, and J. Wu, "Stackelberg game in critical infrastructures from a network science perspective," *Physica A: Statistical Mechanics and its Applications*, vol. 521, pp. 705–714, 2019. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0378437119301268

[8] D. Fudenberg and J. Tirole, *Game Theory*. MIT Press, 2003.

[9] G. Brown, M. Carlyle, J. Salmern, and K. Wood, "Defending critical infrastructure," *Interfaces*, vol. 36, no. 6, pp. 532–544, 2006.

[10] H. Fawzi, P. Tabuada, and S. Diggavi, "Secure estimation and control for cyber-physical systems under adversarial attacks," *IEEE Transactions on Automatic Control*, vol. 59, no. 6, pp. 1454–1467, 2014.

[11] T. G. Lewis, *Critical infrastructure protection in homeland security: defending a networked nation*. John Wiley & Sons, 2014.

[12] F. Pasqualetti, F. Drfler, and F. Bullo, "Attack detection and identification in cyber-physical systems," *IEEE Transactions on Automatic Control*, vol. 58, no. 11, pp. 2715–2729, 2013.

[13] A. Chattopadhyay, A. Prakash, and M. Shafique, "Secure cyber-physical systems: current trends, tools and open research problems," in *Design, Automation & Test in Europe Conference & Exhibition (DATE), 2017*. IEEE, 2017, pp. 1104–1109.

[14] A. Humayed, J. Lin, F. Li, and B. Luo, "Cyber-physical systems securitya survey," *IEEE Internet of Things Journal*, vol. 4, no. 6, pp. 1802–1831, 2017.

[15] V. Gunes, S. Peter, T. Givargis, and F. Vahid, "A survey on concepts, applications, and challenges in cyber-physical systems," *Transactions on internet and information systems*, vol. 8, no. 12, pp. 4242–4268, 2014.

[16] S. C. Suh, U. J. Tanik, J. N. Carbone, and A. Eroglu, Eds., *Applied Cyber-Physical Systems*. Springer-Verlag, 2014.

[17] D. W. McKee, S. J. Clement, J. Almutairi, and J. Xu, "Survey of advances and challenges in intelligent autonomy for distributed cyber-physical systems," *CAAI Transactions on Intelligence Technology*, vol. 3, no. 2, pp. 75–82, 2018.

[18] M. Gerla, E.-K. Lee, G. Pau, and U. Lee, "Internet of vehicles: From intelligent grid to autonomous cars and vehicular clouds," in *2014 IEEE world forum on internet of things (WF-IoT)*. IEEE, 2014, pp. 241–246.

[19] D. Ding, Q.-L. Han, Y. Xiang, X. Ge, and X.-M. Zhang, "A survey on security control and attack detection for industrial cyber-physical systems," *Neurocomputing*, vol. 275, pp. 1674–1683, Jan 2018.

[20] K. Hausken, "Defense and attack for interdependent systems," *European Journal of Operational Research*, vol. 256, pp. 582–591, 2016.

[21] A. A. Cardenas, S. Amin, and S. Sastry, "Secure control: Towards survivable cyber-physical systems," in *The 28th International Conference on Distributed Computing Systems Workshops*. IEEE, 2008, pp. 495–500.

[22] J. Beyza, E. Garcia-Paricio, and J. M. Yusta, "Applying complex network theory to the vulnerability assessment of interdependent energy infrastructures," *Energies*, vol. 12, no. 3, p. 421, 2019.

[23] S. Boccaletti, V. Latora, Y. Moreno, M. Chavez, and D.-U. Hwang, "Complex networks: Structure and dynamics," *Physics reports*, vol. 424, no. 4-5, pp. 175–308, 2006.

[24] T. Alpcan and T. Basar, *Network Security: A Decision and Game Theoretic Approach*. Cambridge University Press, 2011.

[25] N. Ghorbani-Renani, A. D. González, K. Barker, and N. Morshedlou, "Protection-interdiction-restoration: Tri-level optimization for enhancing interdependent network resilience," *Reliability Engineering & System Safety*, vol. 199, p. 106907, 2020.

[26] N. S. V. Rao, Y. Narahari, C. E. V. Madhavan, D. K. Y. Yau, and C. Y. T. Ma, "An analytical framwork for cyber-physical networks," in *Securing Cyber-Physical Infrastructures: Foundations and Challenges*, S. K. Das, K. Kant, and N. Zhang, Eds. Morgan Kaufmann, 2012.

[27] V. M. Bier and M. N. Azaiez, Eds., *Game Theoretic Risk Analysis of Security Threats*. Springer, 2009.

[28] P.-Y. Chen, S.-M. Cheng, and K.-C. Chen, "Smart attacks in smart grid communication networks," *Communications Magazine, IEEE*, vol. 50, no. 8, pp. 24–29, 2012.

[29] M. H. Manshaei, Q. Zhu, T. Alpcan, T. Bacşar, and J.-P. Hubaux, "Game theory meets network security and privacy," *ACM Computing Surveys (CSUR)*, vol. 45, no. 3, p. 25, 2013.

[30] S. Shiva, S. Roy, and D. Dasgupta, "Game theory for cyber security," in *Proceedings of the Sixth Annual Workshop on Cyber Security and Information Intelligence Research*. ACM, 2010, p. 34.

[31] R. Yu, Y. H. Chen, and B. Han, "Cooperative game approach to robust control design for fuzzy dynamical systems," *IEEE Transactions on Cybernetics*, pp. 1–13, 2020.

[32] J. Li, J. Ding, T. Chai, and F. L. Lewis, "Nonzero-sum game reinforcement learning for performance optimization in large-scale industrial processes," *IEEE Transactions on Cybernetics*, vol. 50, no. 9, pp. 4132–4145, 2020.

[33] Q. Zhu and T. Basar, "Game-theoretic methods for robustness, security, and resilience of cyberphysical control systems: Games-in-games principle for optimal cross-layer resilient control systems," *IEEE Control Systems*, vol. 35, no. 1, pp. 46–65, 2015.

[34] A. Ferdowsi, A. Eldosouky, and W. Saad, "Interdependence-aware game-theoretic framework for secure intelligent transportation systems," *IEEE Internet of Things Journal*, 2020.

[35] D. Alderson, G. Brown, W. Carlyle, and R. Wood, "Solving defender-attacker-defender models for infrastructure defense," in *INFORMS Computing Society Conference*, 2011, pp. 28–49.

[36] Y. Li, S. Tan, Y. Deng, and J. Wu, "Attacker-defender game from a network science perspective," *Chaos*, vol. 28, no. 5, pp. 051 102–051 109, 2018.

[37] M. Ouyang, "A mathematical framework to optimize resilience of interdependent critical infrastructure systems under spatially localized attacks," *European J. Oper. Res.*, vol. 262, no. 3, pp. 1072–1084, 2017.

[38] A. Garnaev, M. Baykal-Gursoy, and H. V. Poor, "Security games with unknown adversarial strategies," *IEEE Transactions on Cybernetics*, vol. 46, no. 10, pp. 2291–2299, 2016.

[39] V. R. R. Jose and J. Zhuang, "Technology adoption, accumulation, and competition in multi-period attacker-defender games," *Military Operations Research*, vol. 18, no. 2, pp. 33–47, 2013.

[40] D. Arce, D. Kovenock, and B. Roberson, "Weakest-link attacker-defender games with multiple attack technologies," *Naval Res. Logist.*, vol. 59, no. 6, pp. 457–469, 2012.

[41] F. He and J. Zhuang, "Modelling 'contracts' between a terrorist group and a government in a sequential game," *Journal of the Operational Research Society*, vol. 63, no. 6, pp. 790–809, 2012.

[42] M. Nikoofal and J. Zhuang, "Robust allocation of a defensive budget considering an attacker's private information," *Risk Analysis*, vol. 32, no. 5, pp. 930–943, 2012.

[43] C. Zhang, J. E. Ramirez-Marquez, and J. Wang, "Critical infrastructure protection using secrecy - a discrete simultaneous game," *European Journal of Operational Research*, vol. 242, no. 1, pp. 212–221, 2015.

[44] K. Hausken and F. He, "On the effectiveness of security countermeasures for critical infrastructures," *Risk Analysis*, vol. 36, no. 4, pp. 711–726, 2016.

[45] C. Kiekintveld, M. Jain, J. Tsai, J. Pita, F. Ordez, and M. Tambe, "Computing optimal randomized resource allocations for massive security games," in *Proc. of International Joint Conference on Autonomous Agents and Multiagent Systems*, vol. 2, Jan. 2009, pp. 689–696.

[46] D. Korzhyk, V. Conitzer, and R. Parr, "Complexity of computing optimal stackelberg strategies in security resource allocation games," in *Proc. of the AAAI Conference on Artificial Intelligence*, vol. 24, no. 1, 2010.

[47] Y. Li, S. Qiao, Y. Deng, and J. Wu, "Stackelberg game in critical infrastructures from a network science perspective," *Physica A: Statistical Mechanics and its Applications*, vol. 521, Feb. 2019.

[48] E. C. Paulson, I. Linkov, and J. M. Keisler, "A game theoretic model for resource allocation among countermeasures with multiple attributes," *European Journal of Operational Research*, vol. 252, no. 2, pp. 610–622, 2016.

[49] X. Shan and J. Zhuang, "Hybrid defensive resource allocations in the face of partially strategic attackers in a sequential defender-attacker game," *European Journal of Operational Research*, vol. 228, no. 1, pp. 262–272, 2013.

[50] Y. Mei, W. Han, S. Li, and X. Wu, "A survey of advanced persistent threats attack and defense," in *2021 IEEE Sixth International Conference on Data Science in Cyberspace (DSC)*. Los Alamitos, CA, USA: IEEE Computer Society, oct 2021, pp. 608–613. [Online]. Available: https://doi.ieeecomputersociety.org/10.1109/DSC53577.2021.00096

[51] K. Hausken and G. Levitin, "Review of systems defense and attack models," *International Journal of Performability Engineering*, vol. 8, no. 4, pp. 355–366, 2012.

[52] I. Atzeni, L. G. Ordóñez, G. Scutari, D. P. Palomar, and J. R. Fonollosa, "Noncooperative and cooperative optimization of distributed energy generation and storage in the demand-side of the smart grid," *IEEE transactions on signal processing*, vol. 61, no. 10, pp. 2454–2472, 2013.

[53] T. G. Alghamdi, D. Said, and H. T. Mouftah, "Decentralized game-theoretic scheme for d-evse based on renewable energy in smart cities: A realistic scenario," *IEEE Access*, vol. 8, pp. 48 274–48 284, 2020.

[54] P. Shakarian, H. Lei, and R. Lindelauf, "Power grid defense against malicious cascading failure," in *Proc. of the international conference on Autonomous Agents and Multi-Agent Systems*, 2014, pp. 813–820.

[55] Q. Dai, X. Xu, W. Guo, S. Huang, and D. Filev, "Towards a systematic computational framework for modeling multi-agent decision-making at micro level for smart vehicles in a smart world," *Robotics and Autonomous Systems*, vol. 144, p. 103859, 2021.

[56] R. Aringhieri, M. E. Bruni, S. Khodaparasti, and J. T. van Essen, "Emergency medical services and beyond: Addressing new challenges through a wide literature review," *Computers & Operations Research*, vol. 78, pp. 349–368, 2017.

[57] F. He, S. Chandrasekar, N. S. V. Rao, and C. Y. T. Ma, "Effects of interdependencies on game-theoretic defense of cyber-physical infrastructures," in *International Conference on Information Fusion*, 2019.

[58] N. S. V. Rao, S. W. Poole, C. Y. T. Ma, F. He, J. Zhuang, and D. K. Y. Yau, "Defense of cyber infrastructures against cyber-physical attacks using game-theoretic models," *Risk Analysis*, 2015.

[59] N. S. V. Rao, N. Imam, C. Y. T. Ma, K. Hausken, F. He, and J. Zhuang, "On defense strategies for system of systems using aggregated correlations," in *11th Annual IEEE International Systems Conference*, 2017.

[60] N. S. V. Rao, C. Y. T. Ma, and F. He, "Defense strategies for multi-site cloud computing server infrastructures," in *International Conference on Distributed Computing and Networking*, 2018.

[61] ——, "Defense strategies and expected capacity of high performance computing infrastructures," in *International Conference on Distributed Computing and Networking*, 2019.
[62] K. S. Trivedi and A. Bobbio, *Reliability and Availability Engineering: Modeling, Analysis, and Applications*. Cambridge University Press, Cambridge, UK, 2017.
[63] T. Bedford and R. Cooke, *Probabilistic Risk Analysis: Foundations and Methods*. Cambridge University Press, Cambridge, UK, 2001.
[64] J. T. Blake and K. S. Trivedi, "Reliability analysis of interconnection networks using hierarchical composition," *IEEE transactions on reliability*, vol. 38, no. 1, pp. 111–120, 1989.
[65] L. Lamport, R. Shostak, and M. Pease, "The byzantine generals problem," *ACM Transactions on Programming Languages and Systems*, vol. 4, no. 3, pp. 382–401, July 1982. [Online]. Available: https://www.microsoft.com/en-us/research/publication/byzantine-generals-problem/
[66] D. M. Nicol, W. H. Sanders, and K. S. Trivedi, "Model-based evaluation: from dependability to security," *IEEE Transactions on dependable and secure computing*, vol. 1, no. 1, pp. 48–65, 2004.
[67] B. Schneier, "Attack trees," *Dr. Dobbs journal*, vol. 24, no. 12, pp. 21–29, 1999.
[68] C. Y. T. Ma, D. K. Y. Yau, and N. S. V. Rao, "Markov game analysis for attack-defense of power networks under possible misinformation," *IEEE Transactions on Power Systems*, vol. 28, no. 2, pp. 1676–1886, 2013.
[69] ——, "Scalable solutions of Markov games for smart-grid infrastructure protection," *IEEE Transactions on Smart Grid*, vol. 4, no. 1, pp. 47–55, 2013.
[70] S. Skaperdas, "Contest success functions," *Economic Theory*, vol. 7, no. 2, pp. 283–290, 1996.
[71] B. K. Rai and R. Sarin, "Generalized contest success functions," *Economic Theory*, vol. 40, no. 1, pp. 139–149, 2009.
[72] N. S. V. Rao, C. Y. T. Ma, K. Hausken, F. He, D. K. Y. Yau, and J. Zhuang, "Defense strategies for asymmetric networked systems under composite utilities," in *IEEE International Conference on Multisensor Fusion and Integration for Intelligent Systems*, 2017.
[73] Y. Kim, J. G. Park, and J.-H. Lee, "Security threats in 5g edge computing environments," in *2020 International Conference on Information and Communication Technology Convergence (ICTC)*, 2020, pp. 905–907.
[74] R. Roman, J. Lopez, and M. Mambo, "Mobile edge computing, fog et al.: A survey and analysis of security threats and challenges," *Future Generation Computer Systems*, vol. 78, pp. 680–698, 2018. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0167739X16305635
[75] T. Naughton, S. Hitefield, L. Sorrillo, N. Rao, J. Kohl, W. Elwasif, J. C. Bilheux, H. Bilheux, S. Boehm, J. Kincl, S. Sen, and N. Imam, "Software framework for federated science instruments," in *Proc. of Smoky Mountains Computational Sciences and Engineering Conference*. Springer-Verlag, 2020.
[76] G. N. Sorebo and M. C. Echols, *Smart Grid Security*. CRC Press, 2012.
[77] L. Xiao, N. B. Mandayam, and H. Vincent Poor, "Prospect theoretic analysis of energy exchange among microgrids," *IEEE Transactions on Smart Grid*, vol. 6, no. 1, pp. 63–72, 2015.
[78] F. He, C. J. Agwuegbo, N. S. V. Rao, and C. Y. T. Ma, "A sequential game of defense and attack on an interdependent system of systems," in *International Conference on Information Fusion*, 2018.
[79] F. He, J. Zhuang, and N. S. V. Rao, "Discrete game-theoretic analysis of defense in correlated cyber-physical systems," *Annals of Operations Research*, vol. 294, no. 1, pp. 741–767, 2020.
[80] T. Brown, D. Sarioz, A. Bar-Noy, T. LaPorta, D. Verma, M. Johnson, and H. Rowaihy, "Geometric considerations for distribution of sensors in ad-hoc sensor networks," City University of New York, Tech. Rep., November 2006, Computer Science Technical Reports, TR-2006014.

Nageswara S. V. Rao is a Corporate Fellow at Oak Ridge National Laboratory, where he joined in 1993. He received B. Tech from National Institute of Technology, Warangal, India, M.E. from School of Automation, Indian Institute of Science, Bangalore, India, and PhD in computer science from Louisiana State University. His research areas include high performance and quantum networking, information fusion, machine learning, and federations of science instruments. He is a Fellow of IEEE, and received 2005 IEEE Computer Society Technical Achievement Award and 2014 R&D 100 Award.

Chris Y. T. Ma is an assistant professor at The Hang Seng University of Hong Kong, where he joined in 2016. He received his BEng in Computer Engineering from The Chinese University of Hong Kong; MPhil in Computer Science and Engineering from The Chinese University of Hong Kong; and PhD in Computer Science from Purdue University, West Lafayette, IN, USA. His research interests include performance, privacy, and security study of networks and cyber-physical systems.

Fei He is an associate professor at Texas A&M University-Kingsville, where she joined in 2014. She received PhD in Industrial Engineering from the State University of New York at Buffalo. Her research areas include operations research and game theory with applications in cyber-physical systems and disaster mitigation. She is a member of INFORMS Computing Society and Decision Analysis Society.

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

The public reporting burden for this collection of information is estimated to average 1 hour per response, including the time for reviewing instructions, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing the collection of information. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing the burden, to Department of Defense, Washington Headquarters Services, Directorate for Information Operations and Reports (0704-0188), 1215 Jefferson Davis Highway, Suite 1204, Arlington, VA 22202-4302. Respondents should be aware that notwithstanding any other provision of law, no person shall be subject to any penalty for failing to comply with a collection of information if it does not display a currently valid OMB control number.
PLEASE DO NOT RETURN YOUR FORM TO THE ABOVE ADDRESS.

1. REPORT DATE *(DD-MM-YYYY)*	**2. REPORT TYPE**	**3. DATES COVERED** *(From - To)*

4. TITLE AND SUBTITLE	**5a. CONTRACT NUMBER**
	5b. GRANT NUMBER
	5c. PROGRAM ELEMENT NUMBER
6. AUTHOR(S)	**5d. PROJECT NUMBER**
	5e. TASK NUMBER
	5f. WORK UNIT NUMBER
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)	**8. PERFORMING ORGANIZATION REPORT NUMBER**
9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)	**10. SPONSOR/MONITOR'S ACRONYM(S)**
	11. SPONSOR/MONITOR'S REPORT NUMBER(S)

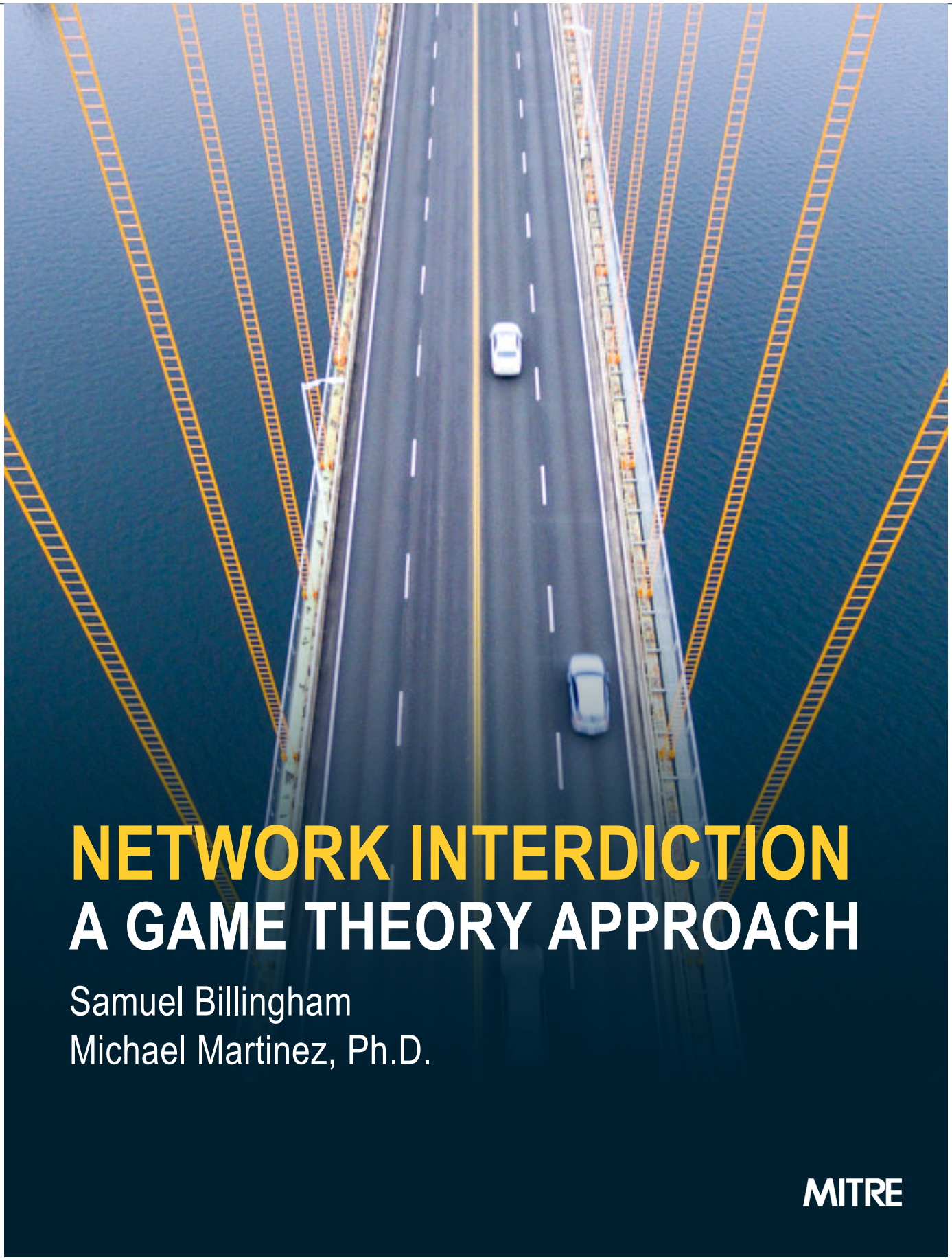
12. DISTRIBUTION/AVAILABILITY STATEMENT

13. SUPPLEMENTARY NOTES

14. ABSTRACT

15. SUBJECT TERMS

16. SECURITY CLASSIFICATION OF:			**17. LIMITATION OF ABSTRACT**	**18. NUMBER OF PAGES**	**19a. NAME OF RESPONSIBLE PERSON**
a. REPORT	**b. ABSTRACT**	**c. THIS PAGE**			**19b. TELEPHONE NUMBER** *(Include area code)*



NETWORK INTERDICTION
A GAME THEORY APPROACH

Samuel Billingham
Michael Martinez, Ph.D.

MITRE

ABSTRACT

Ensuring the resilience of critical infrastructure networks, and adequately defending these networks for the purpose of national security, can be a steep challenge, especially in this age of contested environments. Beyond natural disasters and chance accidents, intelligent adversaries may be capable of inflicting serious operational impact with only a few strategically placed attacks. However, simple priority listings may result in overlooked vulnerabilities that are easily exploited. In these cases, careful analysis is required to identify the critical components (nodes and arcs) of the network, for the purpose of focusing efforts to enhance their resiliency. We discuss a three-player network interdiction construct that combines optimization and game theory, applied to a use case within the U.S. transportation network. This analysis provides a powerful, robust approach to identify critical infrastructure within the network.

©2022 The MITRE Corporation. All Rights Reserved.
Approved for Public Release; Distribution Unlimited. Case 22-1141

THE CHALLENGE

The U.S. government operates and oversees a wide variety of networks, and these systems are often classified as critical infrastructure, essential for the health and welfare of the nation. These networks are generally administered with high efficiency, whether the objective is to maximize flow throughout or to send a specific set of items from "source nodes" to "demand nodes" at minimum cost. However, ensuring resilience of these networks, and adequately *defending* these networks for the purpose of national security, can be a steeper challenge, especially in this age of contested environments. Beyond natural disasters and chance accidents, intelligent adversaries may be capable of inflicting serious operational impact with only a few strategically placed attacks (kinetic and/or non-kinetic). In these cases, careful analysis is required to identify the critical components (nodes and arcs) of the network, for the purpose of focusing efforts to enhance their resiliency.

Generally, security planners in the Department of Defense (DoD) and Department of Homeland Security (DHS) are accustomed to creating a priority listing of network components, highlighting which pieces are most important to protect and defend. Depending on the available resources, the priority determines the set order in which to apply resiliency measures. However, it turns out that such a static priority listing may overlook significant gaps, where entirely different sets of components, if compromised, could combine to have a devastating effect on network operations. Priority lists and simple rules of thumb can be effective, but they do not *guarantee* that such a solution is optimal.

A Counterintuitive Example

Pretend we are charged with defending the simple network shown in **Figure 1**. The goal here is to push the maximum flow through the network, from node A to node F. The values shown on the arcs indicate the capacity of that particular link.

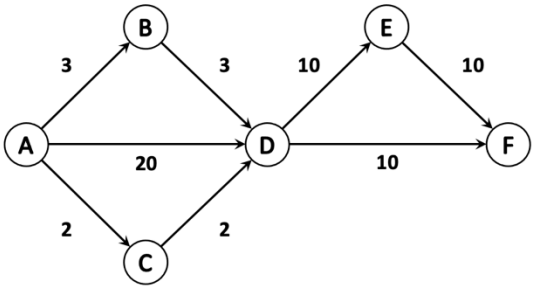


Figure 1. Example Network for Maximum Flow. Source: Naval Postgraduate School

The maximum flow through the network initially is 20 units. An intelligent attacker would clearly snip arc A→D, cutting the maximum flow down to only 5. *So a defender would naturally look to protect that "most-vital" arc.* However, what if the attacker is able to inflict *two* attacks instead of one? In this case, he has two optimal choices: snip D→E and D→F, or snip D→F and E→F. In either case, the flow is reduced completely to zero—and notice that arc A→D is no longer in the discussion! The attacker's strategy changed entirely, when a different budget, or capability "quiver" is considered. This quick and simple example defies a straightforward, prioritized list of critical network components.

Thus more scrutiny must be given when analyzing network flows. *Simple priority listings may result in overlooked vulnerabilities that are easily exploited.* Instead of creating static priority menus, an optimization approach to network interdiction should be adopted.

©2022 The MITRE Corporation. All Rights Reserved.
Approved for Public Release; Distribution Unlimited. Case 22-1141

OUR APPROACH

In our work for combatant commands within the DoD, we employ optimization theory to this problem of network interdiction to ensure the U.S. government (USG) doesn't "overlook" an even stronger, counter-intuitive solution that may be hidden just beneath the obvious vulnerable nodes and links. This classic approach is rooted in mathematics and has been applied to various network flow problems for decades. Indeed, the RAND Corporation studied the Soviet rail system back in the 1950s for network vulnerabilities[1].

Many times, optimizing network flows is actually a simple problem, allowing most operators to run their systems at maximum efficiency. However, in light of hostile intentions, the situation becomes much more complex and the solutions not so obvious. In order to account for both sides, the optimization can be combined with Game Theory—specifically, the Stackelberg model. In this paradigm, the players take turns sequentially and have full, open knowledge of each other's moves, as in chess. The players' objectives are diametrically opposed, each seeking to counter the other's moves. Combining both optimization and game theory in this fashion provides a powerful approach to this class of network planning and interdiction problems.

The Naval Postgraduate School has researched these "Attacker-Operator" designs for over ten years, and has developed a robust library of applied models[2]. In fact, they have extended the construct to include a third player, a *Defender*, who seeks to assist the Operator by emplacing defenses at various components of the network. This interdiction game can be termed "Defender-Attacker-Operator" since that is the order in which the moves occur. As we saw with the sample network, critical network components are not so easily identified, so adding a third player allows for a much more rigorous cross-check of network performance.

COMBINING BOTH OPTIMIZATION AND GAME THEORY IN THIS FASHION PROVIDES A POWERFUL APPROACH TO THIS CLASS OF NETWORK PLANNING AND INTERDICTION PROBLEMS.

A Solution Methodology

Moving first but possessing a limited budget, the Defender decides where to place defenses within the network. Seeing where these defenses are laid, the Attacker next chooses which components to attack, likely some combination of arcs and nodes. (One assumption here is that the Attacker will avoid defended components entirely and instead concentrate his strikes elsewhere.) Finally, the Operator must flow commodities across this network that has now been both defended and attacked—considering that defended components work as intended, while attacked components experience some measure of degradation, or perhaps temporary or entire failure.

As when exploring for a Nash Equilibrium within Game Theory, the players take turns optimizing from their perspective, while the resulting network performance metrics begin to converge. If minimum time (shortest path) is the primary objective for the game, then the Defender and Operator are choosing defenses and routes that will allow for

[1] See Alderson *et al.,* 2013.

[2] See Alderson *et al.,* 2011.

©2022 The MITRE Corporation. All Rights Reserved.
Approved for Public Release; Distribution Unlimited. Case 22-1141

minimum overall transit time of the "cargo" in the network. The Attacker, however, intends to delay as much cargo as possible by as long as possible. As the game iterates, the players "learn" how to counter the opposing side, and may find better and better strategies to respond to their opponent while maintaining strong performance on their part. Eventually, the performance metrics balance, and no player can improve their strategy without sacrificing performance elsewhere—the solutions have converged.

With three players, the solution algorithm is fairly complicated, requiring two separate loops and passing solutions between players. The algorithm is outlined in **Figure 2**. Note that at each stage of the game, the current solution for various players is "locked" into place, while the opponent seeks the optimum strategy to counter the current status of the board.

In reality, the Defender moves first, followed by the Attacker and then the

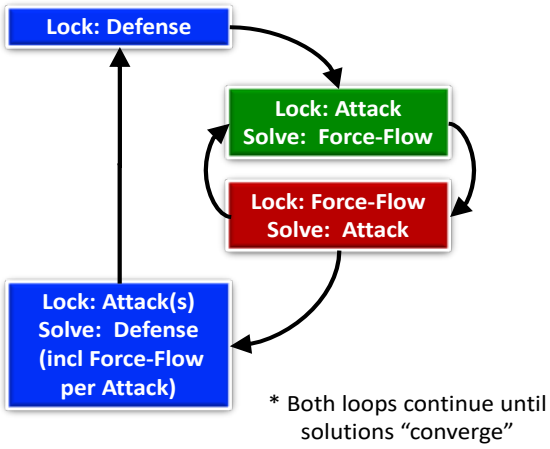


Figure 2. Solution Algorithm
Source: MITRE

Operator. However, the solution algorithm works in reverse, having the Operator and the Attacker engage first, seen in the green and red procedures; this inner loop iterates until their two solutions converge. At that point, the resulting attack plan is passed to and locked into place by the Defender, who tests all possible defenses against it, while simultaneously examining the subsequent Operator flow plan that would accompany each possible defense option. Once that solution is determined, that defense plan is locked into place and passed back to the inner loop, where the Attacker and Operator again iterate. As the attack plans are resolved, these are collected into a list, so the Defender must account for each of them—which places him in a "worst-case" scenario, unsure which attack plan the Attacker may utilize. This outer loop must converge in order for the overall solution to be decided. At that point, the solution yields an ideal defense strategy, a "worst-case" attack plan, and a corresponding optimal flow routing to mitigate the attack effects, while likely making use of defended infrastructure.

USE CASE

In 2019, NORAD & USNORTHCOM tasked MITRE to conduct a Resiliency Analysis of infrastructure critical to the force projection mission within the Continental United States (CONUS). The effort sought to identify transportation network assets critical to timely force projection[3] of military equipment and ammunition from bases through Ports of Embarkation over three transportation modes: rail, road, and oceanic shipping. The effort was limited to identifying only strategic infrastructure (roads, rail lines, intersections, ports, and shipping lanes), so it does not consider individual transportation vehicles (trucks, locomotives, railcars, vessels).

[3] A unit's expected arrival date in theater determined based on theater mission objectives.

©2022 The MITRE Corporation. All Rights Reserved.
Approved for Public Release; Distribution Unlimited. Case 22-1141

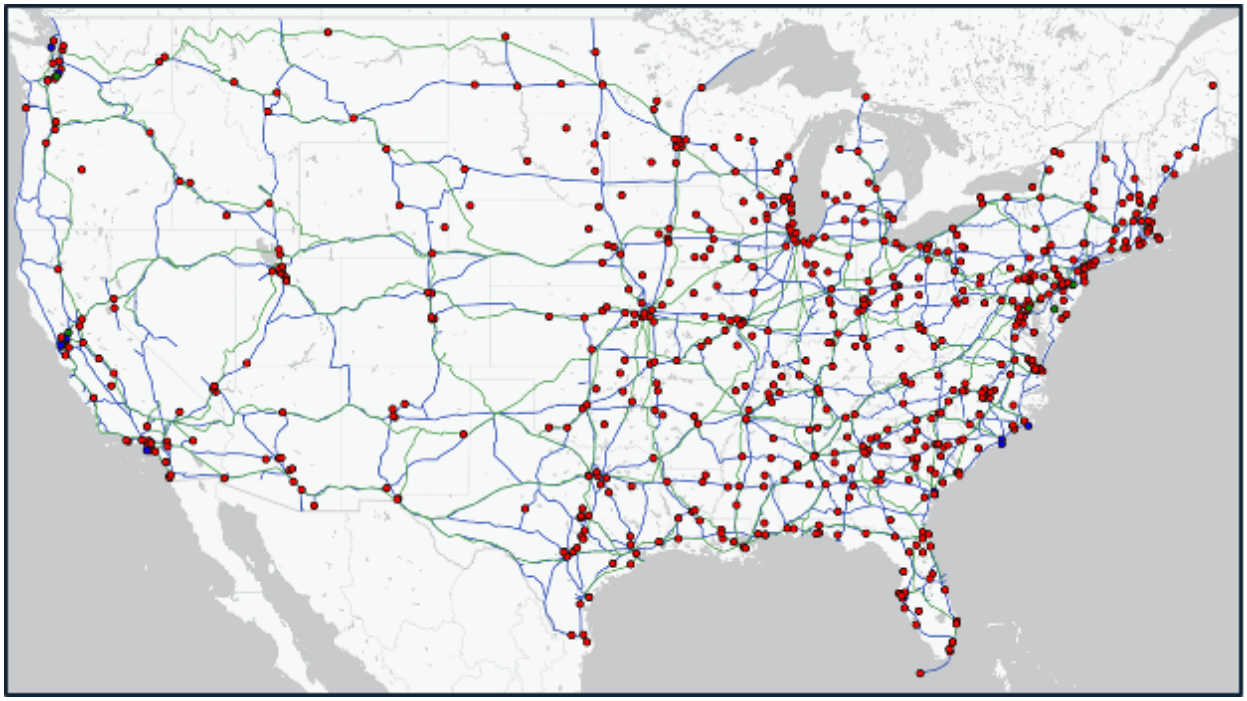


Figure 3. CONUS Transportation System with origins and ports (road - blue, rail - green)
Source: MITRE

We began with the network infrastructure. **Figure 3** shows the designated road and rail networks used to move equipment and ammunition throughout CONUS, as well as the origins of the equipment and ammunition. This network contains approximately 3,800 nodes and 5,200 arcs. To identify critical infrastructure, we assumed a minimum time objective, where an operator would flow network "cargo" items in as timely a manner as possible, an adversary injects delays by attacking the infrastructure, and a defender attempts to thwart these delaying attacks by defending the infrastructure. Along with the network, four types of data are required. First, the Time Phased Force Deployment Data (TPFDD) for an Operational Plan (OPLAN) provides the locations and amounts (in short tons) of planned equipment and ammunition to be deployed. Approximately 2,000+ unique items are modeled and moved across the infrastructure. Second, transit times (in hours) are needed along the various arcs (rail, road, shipping links) in the network. Third, processing times (in short tons per hour) are required for the various handling nodes within the network. Lastly, expected delay times for arcs (in hours) or nodes (in short tons per hour), should they be attacked, are required.

A model based on the Defender-Attacker-Operator (DAO) construct was used to identify critical infrastructure under various attacker and defender budget combinations along with flow times for the cargo. While detailed results of our analysis are sensitive, we provided a list of critical infrastructure to NORAD & USNORTHCOM, as well as details of what cargo is being delayed and the amount of delay under various attacker/defender budget (and thus, capability) combinations. Note that adjusting the budget combinations allowed us to identify any fundamental changes to the Attacker or Defender strategies, as shown in the initial example.

Infrastructure that is defended consistently across budget scenarios is considered more

©2022 The MITRE Corporation. All Rights Reserved.
Approved for Public Release; Distribution Unlimited. Case 22-1141

critical to timely force projection than infrastructure that is routinely attacked, which in turn is considered more critical than infrastructure that is typically neglected. The critical infrastructure lists allow for focusing of resources to conduct in-depth vulnerablity assessments at the identified infrastructure. Also, theater delay results support mission risk assessments due to late arrivals in the receiving theater. In general, we found that an attacker can inflict enough delay into the force projection network to place theater mission objectives at risk with very small budgets. We also found (to our relief!) that mitigating much of the delay is possible with the application of even fewer defenses.

ADVANTAGES & DISADVANTAGES

The advantage of the DAO methodology is its extensibility. While the use case demonstrated its applicability to a transportation network, it is not limited to transportation. As long as the underlying system is a network (e.g, communications, command & control) the methodology can be used to identify where potentially dangerous vulnerabilities exist in the system. Likewise, while the use case demonstrates a minimum-timeliness example, the methodology also supports other (minimum-cost, maximum-flow) objectives.

Another advantage of this methodology is its practicality. Because it is simply searching for elements where vulnerabilites exist, and not describing these vulnerabilites at an engineering level, it allows decision makers, via a broad look at a complex system, to readily target where risks to their missions could exist when facing an intelligent adversary. It also allows for the quick identification of where they need to focus future efforts and resources for risk analyses and vulnerability assements. Also, while not demonstrated in the use case, the method could be adapted for two practical extensions: distinctions between attack types (e.g. kinetic versus non-kinetic attacks) and the impacts of area attacks/defenses (i.e., attacks/defenses that, if employed, affect multiple components of the network simultaneously). While this adds some complexity to the analysis, it is not insurmountable.

Another advantage of this approach is that it is not as sensitive to erroneous input data as others. That is, utilizing the best available data, one can readily identify vulnerabilities.

WHILE NO METHODOLOGY COMPLETELY SURVIVES "GARBAGE IN," THIS METHOD IS FAIRLY ROBUST TO PERTURBATIONS IN THE INPUT DATA.

Much of the input data used for the handling and movement times in the use case can be found in open sources. Likewise, since attack impacts are less known, subject matter expert estimates were used and applied consistently across the network; vulnerable nodes are still identified. Sensitivity analyses for these rough estimates were conducted, showing that the results hold up to moderate changes in the original inputs. If leaders (or analysts) are concerned that the input data is not precise enough to trust the initial results, then the inputs may still be used to identify and narrow the focus for follow-on data collection efforts, streamlining the refinement task.

A couple of disadvantages of this method are the requirement of specialized software and high-powered computing to solve these problems on a massive scale such as our

©2022 The MITRE Corporation. All Rights Reserved.
Approved for Public Release; Distribution Unlimited. Case 22-1141

example. While optimization solvers range anywhere from the simple engine found in MS Excel to powerful commercial solvers (e.g., Gurobi) and anywhere in between, the scale of these problems often drives the need for more expensive options. There has been advancement in open-source solvers over the years and some could be used to solve smaller versions of these network problems. The computing challenge centers around the need for a capable, single machine to solve the problem. That is, the "solve" operation for these DAO problems is not parallelizable, the way most modern simulation efforts or gaming applications are. Thus, a single powerful processor is needed. Luckily for our use case, we were able to utilize an on-hand computer at NORAD & USNORTHCOM.

CONCLUDING THOUGHTS

Network interdiction is becoming an important aspect of national security planning, since both conventional and (oftentimes simplistic) asymmetric threats can be readily employed to cripple critical networks, for discrete windows of time. For example, the crash of the Hong Kong-flagged ship, the *Ever Given*, dramatically affected international shipping by blocking the Suez Canal for six days in 2021. Our "Forts-to-Ports" application allowed security planners to focus future resiliency enhancing efforts, and requests for resources where they are needed most, by readily identifying critical network components. By knowing where to look, the effects of intelligent attacks can more readily, rapidly, and to the maximum extent possible, be mitigated.

ABOUT THE AUTHORS

Samuel Billingham is an Operations Research Analyst with The MITRE Corporation. He holds a M.S. in Operations Research from The University of Alabama in Huntsville.

Michael Martinez is an Operations Research Analyst with The MITRE Corporation. He holds a Ph.D. in Operations Research from the Colorado School of Mines, and served on active duty with the U.S. Air Force before retiring as a Lieutenant Colonel in 2020.

ABOUT MITRE

MITRE's mission-driven teams are dedicated to solving problems for a safer world. Through our public-private partnerships and federally funded R&D centers, we work across government and in partnership with industry to tackle challenges to the safety, stability, and well-being of our nation.

The views, opinions, and/or findings contained herein are those of the authors and should not be construed as an official government position, policy, or decision unless designated by other documentation.

©2022 The MITRE Corporation. All Rights Reserved.
Approved for Public Release; Distribution Unlimited. Case 22-1141

REFERENCES

Alderson, D. L., Brown, G. G., Carlyle, W. M., & Wood, R. K. (2011). *Solving Defender-Attacker-Defender Models for Infrastructure Defense*. in ICS 2011, 12th INFORMS Computing Society Conference, pp 28–49. https://doi.org/10.1287/ics.2011.0047

Alderson, D. L., Brown, G. G., Carlyle, W. M., & Cox, L. A. (2013). *Sometimes There is No 'Most-Vital' Arc: Assessing & Improving the Operational Resilience of Systems.* Military Operations Research, Vol. 18 No. 1, pp 21–37. doi: 10.5711/1082598318121

©2022 The MITRE Corporation. All Rights Reserved.
Approved for Public Release; Distribution Unlimited. Case 22-1141

arXiv:2111.01876v1 [cs.GT] 2 Nov 2021

Article

Game Theory in defence applications: a review

Edwin Ho[1], Arvind Rajagopalan[2], Alex Skvortsov[2], Sanjeev Arulampalam[2], and Mahendra Piraveenan[1,*]

[1] Faculty of Engineering, University of Sydney, Sydney, NSW 2006, Australia
[2] Maritime Division, Defence Science and Technology (DST) Group, Australia
* Correspondence: mahendrarajah.piraveenan@sydney.edu.au

Received: date; Accepted: date; Published: date

Abstract: This paper presents a succinct review of attempts in the literature to use game theory to model decision making scenarios relevant to defence applications. Game theory has been proven as a very effective tool in modelling decision making processes of intelligent agents, entities, and players. It has been used to model scenarios from diverse fields such as economics, evolutionary biology, and computer science. In defence applications, there is often a need to model and predict actions of hostile actors, and players who try to evade or out-smart each other. Modelling how the actions of competitive players shape the decision making of each other is the forte of game theory. In past decades, there have been several studies which applied different branches of game theory to model a range of defence-related scenarios. This paper provides a structured review of such attempts, and classifies existing literature in terms of the kind of warfare modelled, the types of game used, and the players involved. in terms of the warfares modelled, we recognise that most papers which apply game theory in defence settings are concerned with command and control warfare, and can be further classified into papers dealing with (i) Resource Allocation Warfare (ii) Information Warfare (iii) Weapons Control Warfare (iv) Adversary Monitoring Warfare. We also observe that most of the reviewed papers are concerned with sensing, tracking, and large sensor networks, and the studied problems have parallels in sensor network analysis in the civilian domain. In terms of the games used, we classify the reviewed papers into papers that use non-cooperative or cooperative games, simultaneous or sequential games, discrete or continuous games, and non-zero sum or zero-sum games. Similarly, papers are also classified into two-player, three-player or multi-player game based papers. We also explore the nature of players and the construction of payoff functions in each scenario. Finally, we also identify gaps in literature where game theory could be fruitfully applied in scenarios hitherto unexplored using game theory. The presented analysis provides a concise summary about the state-of-the-art with regards to the use of game theory in defence applications, and highlights the benefits and limitations of game theory in the considered scenarios.

Keywords: Decision making; Game theory; Defence Science; Ground warfare; Maritime warfare; Tracking

1. Introduction

Game Theory has become one of the conventional theoretical frameworks to model important decision making processes in many aspects of our life. The well-known examples include economy, social sciences, finance, population dynamics and epidemics (see [1–5] and references therein). Since the seminal work of John Von Neumann, John Nash, and others [6–8], it has been well recognised that there is an optimal strategy in the context of complex interactions (games) between two or more parties (players) that can lead to a predictable outcome (payoff). In practical situations this outcome can often be quantitative and amenable to arithmetic operations (cost, number of infected people, number of vaccinated people etc), but often it can be qualitative in nature (such as risk, readiness level, health state etc).

Application of game theory in defence has a long and diversified history ranging from the design of real time military systems (e.g, applied in missile interception) to the support of strategic decisions on large defence investments and acquisitions. There is extensive literature on specific theoretical methods and tools and their defence applications [9–12]. We believe that the review of this literature is of interest to the community dealing with operational analysis and data-driven decision support. This is the main motivation for the presented study.

Game theory [13,14] elevates military strategies and decisions with a more holistic and value-laden analysis of the situation [9]. For the military, the potential scenarios amenable to game theoretic analysis include the rapidly growing applications of autonomous intelligent systems, and game theory provides a comprehensive mathematical framework that greatly enhances the capabilities of decision making of the people who use these systems. Because of its potential, research into game theory is burgeoning, with a few papers beginning to emerge in the literature for this military research niche. The aim of this review is to assist researchers in utilising the body of knowledge in game theory to develop smarter and safer software systems for defence practitioners. Given that the state of such research is still in an incipient phase, we do this by drawing connections between existing military knowledge with the nascent possibilities that game theory offers so that it can become a more widely understood and considered framework in military software systems.

Although it is not overly extensive, the body of literature around game theory in the military has covered a notable portion of the different forms of engagement and combat. These papers cover past, present and future scenarios: from predictive strategies in potential hostile situations to analytical assessments in hindsight of military standoffs thousands of years ago. Game theory has demonstrated the capacity to be useful in any such military scenario. However, rapid technological progression has led to consistently new frontiers of military engagement, each of these possessing its own complex systems. The overarching areas that have been addressed are Tracking systems (across all domains), Aerial combat, Ground combat, Homefront warfare, Cyber warfare and Space systems. Notably, applications of game theory in Naval warfare have been few, and an exploration into the future research into areas like this will be discussed later in the review. Within each of these areas there are a myriad of possibilities for new and innovative systems: different agents, different weapons, different control structures - and each of these could be enriched with game theoretic analysis. While Haywood's and Thunholm's treatises on game theory used in military decision making provide coverage of several different game types [15,16], there does not seem to be a paper which addresses the use of game theory in the military across each of the respective fields in the new context of military systems built on complex and high performance processing and algorithms. Our aim is to present the literature in such a way that it addresses all of the functions of game theory in military software systems in each key domain.

This review highlights the scope and utility of each analysed paper by presenting it in terms of the essential game components: players, game types, strategies and the key parameters of their payoff functions. It will act as both an annotated bibliography as well as a framework to understand and plan further research into the area. It will also layout the fundamental tenets that are considered by players in every military decision making scenario, as well as how they impact the decisions that are made by military personnel and systems, either while competing with hostile players or while cooperating with friendly players. This will make it possible for most military scenarios to be viewed as games, and can provide, at the very least, an interesting new perspective on familiar military situations.

The rest of the paper is structured as follows. Section 2 will discuss the basic defence principles which are elaborated by the papers that we review, as well as introduce basic concepts of game theory. Section 3 investigates and analyses the literature and summarises the findings and associations in each of the papers. Section 4 elaborates our multi-dimensional classification of the literature based on the observations made in the previous section. Section 5 identifies the gaps in the literature, and based on this, highlights

opportunities for future research in niche, particularly areas of defence research which could benefit from the application of game theory where game theory has not been often applied so far. Finally, section 6 summarises our findings and classifications, and provides broad conclusions.

2. Background

Ideologies, beliefs and knowledge about war have been shaping human knowledge and philosophy for centuries. The great works of Sun Tzu, Homer and Machiavelli [10–12] have not only established a foundation for knowledge etched into the essence of military decision making, but also provided insight into sociology and social psychology [17]. The military forms a core power bloc for many civilisations, and is instrumental to both the growth of influence for existing nations, and the birth of new nations [18]. The military deals with conflicts in real time, plans for the future, as well as reviews past engagements - and every single one of these activities has an impact on society [19]. This review therefore by necessity addresses many facets of military conflict across multiple physical domains, and the major decisions which define each domain will be further clarified below. With respect to game theory for all of these domains, the value of targets, the value of resources, and the priority of objectives are usually the key parameters that shape the payoff functions and strategies which in turn define the games that we use in modelling.

In this section, we discuss the concepts in defence science and technology, as well as game theory, which are necessary to understand and analyse the literature in the presented niche. First of all, let as consider the broad domains of defence and national security which are considered in this review. They can be summarised, as shown in Table 1.

Table 1. Classification System used in this review

Military's Main Method	Command and Control Warfare				
Period	Traditional (T)			Modern (M)	
Military Field	Land (L)	Air (A)	Sea (S)	Cyber (C)	Space (S)
Major Military Field	Resource Allocation Warfare (RAW) or Information Warfare (IW) or Weapon Control Warfare (WCW) or Adversary Monitoring Warfare (AMW)				
Game Theory Category 1	Non-Cooperative (NCo) or Cooperative (Co)				
Game Theory Category 2	Sequential (Seq) or Simultaneous (Sim)				
Game Theory Category 3	Discrete (D) or Continuous (C)				
Game Theory Category 4	Zero Sum (ZS) or Non-Zero Sum (NZS)				

Table 1. Classification System used in this review

Military's Main Method	Command and Control Warfare
Number of Players	2 player (2P) or 3-player (3P) or N-player (NP)

As shown in the table 1, in this review, the focus is primarily on 'command and control' warfare, where decision making is critical. However, command and control warfare has applicability in traditional modes of warfare, such as land, sea, and air warfare, as well as modern modes of warfare, such as space and cyber warfare. At an orthogonal level, command and control warfare could also be sub-divided into Resource Allocation Warfare (RAW), Information Warfare (IW), Weapons Control Warfare (WCW), as well as Adversary Monitoring Warfare (AMW). Since these concepts are extensively used in our classification of literature, let us briefly introduce them first.

2.1. Warfare types

2.1.1. Land warfare

Technology is a dictating force of warfare, and technology is not as imperative to land warfare as it is to other domains[20]. The technology that has impacted land warfare has been static, and avoids the exposure of human resources if possible [21,22]. Interpersonal combat at a physical level is much less prevalent nowadays, making way for a greater focus on positioning strategy. The literature which applied game theory to ground warfare includes a strong repository of Weapon-Target Allocation papers (which touch upon Weapon Control Warfare and Resource Allocation Warfare in the modern context), as well as papers that address ancient ground engagements and guerilla warfare. Where human lives are vulnerable, their protection is the most important element of these games, and the next priority is the protection of ground based assets.

2.1.2. Sea (Naval) warfare

Given the importance of navies for the projection of power globally, there is a surprising paucity of publicly available literature on naval warfare - with or without the application of game theory. There is often mention of naval warfare in papers dealing with target tracking, but a discussion on military naval strategy is limited to outdated literature or discussion of bare essentials [23,24]. We will review the available papers in this regard, and highlight this an area where there is a sizeable gap in the literature.

2.1.3. Aerial warfare

It was not long after the Wright Brothers' invention of the aeroplane that aerial warfare became a critical factor in combat and military campaigns [25]. In a combat medium rarely impeded by obstacles or dimension, the nature of aerial combat is fast paced, intuitive and incredibly treacherous, with unpredictable 'rules' for engagement [26,27]. In the present day, the factors to consider are vastly more complicated, and there is no shortage of resources available to military forces to conduct aerial combat-both human and machine [28–30]. The literature shows that as a result of this abundance of arsenal, the intrinsic and potential value of both the targets and the resources used to engage is of particular importance in aerial warfare scenarios. Decisions about those values for both sides of the conflict need to be made

when evaluating strategies for combat. As such, there are several papers which deal with the use of game theory in aerial warfare.

2.1.4. Space warfare

While the notion of warfare in space has existed for almost a century, neither physical execution nor established strategic theory of space warfare has been established [31]. Nevertheless, this has not stopped military forces chasing the stars (literally and figuratively) [32,33], and has inevitably lead to concepts from game theory being used in space warfare strategic thinking. This is currently mostly limited to satellite networks, where the key parameters of the game are optimised power use and signal strength across the network. The field is still quite young, and further military development in space seems to be inevitable, with which the corresponding literature dealing with applications of game theory in space warfare will also grow.

2.1.5. Cyber warfare

Cybersecurity is the protection of IT systems and networks from being damaged/disrupted/subjected to information theft. Cyber warfare deals with the concept of Information and communication Systems being deliberately attacked to obtain a military advantage. While Cybersecurity has been an important field in computer science for many decades now, literature about cyber warfare as such is more scarce, and in any case heavily overlaps with applications of game theory in computer science in areas related to cybersecurity. This review presents and analyses some papers which are specifically concerned with cyber warfare.

2.1.6. Mixed/ other warfare

There are several papers which address specific niches of defence applications of game theory, and yet cannot be classified as papers analysing a certain type of warfare. In some of these papers, the focus is more on the technology that is used: for example, target tracking. In others, the nature of the hostile actors against whom the defence needs to be conducted changes: for example, national security operations which target domestic terrorism threats. There are several papers which deal with the use of game theory in such scenarios.

Target tracking systems: Target tracking in the military is the observation of a moving target and the surveillance of its position and manoeuvres [34–36]. Success in this domain relies on accuracy in the observed metrics and data, as well as efficient distribution and processing of all collected information [37]. With the advent of intelligent targets, the military must also incorporate predictive methods to maintain ideal tracking performance. The literature reviewed in tis regard covers topics from tracking strike missiles to theatre ballistic missiles, and tracking unknown intelligent agents to enemy aircraft. Key considerations in this area which shape the games played involve whether or not the target is 'intelligent' / can take evasive action, weather or not target will have an optimal trajectory, and whether or not the target will have defenders [38]. Target tracking applications of game theory mostly occur in aerial and naval warfare, including underwater surveillance.

National Security applications : Game theory often finds application in national security and anti-terrorism related applications. This includes predicting and preparing for terrorist attacks, as well as resource allocation scenarios for protection of key personnel and landmarks / other potential targets for terrorist activity. While the value of potential targets and the likelihood of attacks are obviously key parameters governing the payoff functions of games in this niche, the subsequent social, economic and political ramifications are equally instrumental in modelling games in this area [39,40]. Few military conflicts have as much exposure as those on the home front [41], and the fallout from terrorist attacks

and their effect on public mood and confidence in the security apparatus are often taken into account in modelling payoff functions in this area.

2.2. Game Theory

Game theory, which is the study of strategic decision making, was first developed as a branch of micro-economics [6,14,42–46]. However, later it has been adopted in diverse fields of study, such as evolutionary biology, sociology, psychology, political science, project management, financial management and computer science [1–5,9,13,47–51]. Game theory has gained such wide applicability due to the prevalence of strategic decision making scenarios across different disciplines. Game Theory provides insight into peculiar behavioural interactions like the cooperative interactions within groups of animals [52,53], the bargaining and exchange in a marriage [54–56] or the incentivisation of Scottish salmon farmers [57]. A typical game defined in game theory has two or more players, a set of strategies available to these players, and a corresponding set of payoff values (sometimes called utility values) for each player (which are, in the case of two-player games, often presented as a payoff-matrix)[47,58,59].

2.2.1. Pure vs Mixed Strategies

A pure strategy in a game provides a complete definition of how a player will play a game. A player's strategy set is the set of pure strategies available to that player[60].

A mixed strategy is a combination of pure strategies where a particular probability p (where $0 \leq p \leq 1$) is associated with each of these pure strategies. Since probabilities are continuous, there are infinitely many mixed strategies available to a player. A totally mixed strategy is a mixed strategy in which the player assigns a strictly positive probability to every pure strategy. Therefore, any pure strategy is actually a degenerate case of a mixed strategy, in which that particular strategy is selected with probability 1 and every other strategy is selected with probability 0. A totally mixed strategy is a mixed strategy in which every pure strategy has a strictly positive (non-zero) value.

2.2.2. Nash Equilibrium

Nash equilibrium is one of the core concepts in game theory. It is a state (a set of strategies) in a strategic game from which no player has an incentive to deviate unilaterally, in terms of payoffs. Both pure strategy and mixed strategy Nash equilibria can be defined. A strategic game can have more than one Nash equilibrium [7,61]. It is proven that every game with a finite number of players in which each player can choose from finitely many pure strategies has at least one Nash equilibrium [7].

The formal definition of Nash equilibrium is as follows. Let (S, f) be a game with n players, where S_i is the strategy set of a given player i. Thus, the strategy profile S consisting of the strategy sets of all players would be, $S = S_1 \text{ x } S_2 \text{ x } S_3 .... \text{ x } S_n$. Let $f(x) = (f_1(x), ..., f_n(x))$ be the payoff function for strategy set $x \in S$. Suppose x_i is the strategy of player i and x_{-i} is the strategy set of all players except player i. Thus, when each player $i \in 1, ..., n$ chooses strategy x_i that would result in the strategy set $x = (x_1, ..., x_n)$, giving a payoff of $f_i(x)$ to that particular player, which depends on both the strategy chosen by that player (x_i) and the strategies chosen by other players (x_{-i}). A strategy set $x^* \in S$ is in Nash equilibrium if no unilateral deviation in strategy by any single player would return a higher utility for that particular player [62]. Formally put, x^* is in Nash equilibrium if and only if:

$$\forall i, x_i \in S_i : f_i(x_i^*, x_{-i}^*) \geq f_i(x_i, x_{-i}^*) \tag{1}$$

2.2.3. Normal-form games and extensive-form games

In a normal-form game, only a single round of decision making takes place, where all players make decisions simultaneously. An extensive-form game is, on the other hand, an iterative game where there are several rounds of decision making [63,64]. On each round, players could make decisions simultaneously or in some pre-defined order. An extensive-form game is often represented by a game tree, where each node (except terminal nodes) is a decision point, and each link corresponds to a decision or a set of decisions that could be made by the relevant player / players at that point. The terminal nodes represent an end to the extensive-form game, with corresponding payoffs for each player involved.

2.2.4. Non-cooperative games and cooperative games

Typically, games are taken to be played for the self-interest of the players, and even when the players cooperate, that is because cooperation is the best strategy under the circumstances to maximise the individual payoffs of the players. In such games, the cooperative behaviour, if emerges, is driven by selfish goals and is transient. These games can be termed 'non-cooperative games'. These are sometimes referred to, rather inaccurately, as 'competitive games'. Non-cooperative game theory is the branch of game theory that analyses such games. On the other hand, in a cooperative game, sometimes also called a coalitional game, players form coalitions, or groups, often due to external enforcement of cooperative behaviour, and competition is between these coalitions [7,8,13,61,65]. Cooperative games are analysed using cooperative game theory, which predicts which coalitions will form and the payoffs of these coalitions. Cooperative game theory focusses on surplus or profit sharing among the coalition [66], where the coalition is guaranteed a certain amount of payoff by virtue of the coalition being formed. Often, the outcome of a cooperative game played in a system is equivalent to the result of a constrained optimisation process [67], and as such, some papers that we review use a linear programming framework to solve the cooperative games they model.

2.2.5. Zero-sum games

Zero-sum games are a class of competitive games where the total of the payoffs of all players is zero. In two player games, this implies that one player's loss in payoff is equal to another player's gain in payoff. A two player zero-sum game can therefore be represented by a payoff matrix which shows only the payoffs of one player. Zero-sum games can be solved with the mini-max theorem [68], which states that in a zero-sum game there is a set of strategies which minimises the maximum losses (or maximises the minimum payoff) of each player. This solution is sometimes referred to as a 'pure saddle point'. It can be argued that the stock market is a zero-sum game. In contrast, most valid economic transactions are non zero-sum since each party considers that, what it receives is more valuable (to itself) than what it parts with.

2.2.6. Perfect vs Imperfect Information Games

In a perfect information game, each player is aware of the full history of the previous actions of all other players, as well as the initial state of the game. In imperfect information games, some or all players do not have access to the entirety of information about other players' previous actions.

2.2.7. Simultaneous games and sequential games

A simultaneous game is either a normal-form game or an extensive-form game where on each iteration, all players make decisions simultaneously. Therefore, each player is forced to make the decision without knowing about the decisions made by other players (on that iteration). On the contrary, a sequential game

is a type of extensive-form game where players make their decisions (or choose their strategies) in some predefined order [63,64]. For example, a negotiation process can be modelled as a sequential game if one party always has the privilege of making the first offer, and the other parties make their offers or counteroffers after that. In a sequential game, at least some players can observe at least some of the actions of other players before making their own decisions (otherwise, the game becomes a simultaneous game, even if the moves of players do not happen simultaneously in time). However, it is not a must that every move of every previous player should be observable to a given player. If a player can observe every move of every previous player, such a sequential game is known to have 'perfect information'. Otherwise the game is known to have 'imperfect information'. Sequential games are often used by papers that we are reviewing here, to model bargaining or negotiation processes.

2.2.8. Differential Games

Differential Games are often extensive form games, but instead of having discrete decision points, they are modeled over a continuous time frame [69]. In such games, the each state variable evolves continuously over time according to a differential equation. Such games are ideal for modelling rapidly evolving defense scenarios where each player engages in selfish optimisation of some parameter. For example, in missile tracking problems, the pursuer and the target both try to control the distance between them, whereas the pursuer constantly tries to minimise this distance and the target constantly tries to increase it. In such a scenario, iterative rounds of decision making are much too discrete to model the continuous movements and computations of each player. Differential games are ideal to model such scenarios.

2.2.9. Stackelberg games

A Stackelberg game is a particular type of two player sequential game commonly used in economics [70]. In a Stackelberg game, there is a leader and a follower, which are typically companies operating in a market. The leader firm has some sort of market advantage that enables it to move first and make the first decision, and the follower's optimal decision depends on the leader's decision. If a follower chose a non-optimal action given the action of the leader, it will adversely affect not only the payoff of the follower, but the payoff of the leader as well. Therefore, the optimal decision of the leader is made on the assumption that the follower is able to see the leader's action and will act to maximise its own payoff given the leader's action. Several papers we review here have used Stackelberg games to model project management scenarios. Stackleberg games might be relevant in defence applications where there are leaders and followers, such as naval, airplane fighter or tank formations.

2.2.10. Common interest games

Common interest games are another class of non-cooperative games in which there is an action profile that all players strictly prefer over all other profiles [71]. In other words, in common interest games, the interests of players are perfectly aligned. It can be argued that common interest games are the antithesis of zero-sum games, in which the interests of the players are perfectly opposed so that any increase in fortune for one player must necessarily result in the collective decrease in fortune for others. Common interest games were first studied in the context of cold war politics, to understand and prescribe strategies for handling international relations [72–74].

It makes sense to classify non-cooperative games into common interest games and non-common interest games, just as much as it makes sense to classify them into zero-sum games and non-zero sum games, since these two concepts (zero-sum games and common interest games) represent extreme cases of non-cooperative games. However, the papers that we review do not use common interest games to model

project management scenarios, and indeed it would be rare to find scenarios in project management where the interests of players are perfectly aligned. Therefore, we do not use the common interest games-based classification in our classification process, as it would add another dimension and increase the complexity of classification needlessly.

2.2.11. Nash bargaining

In a Nash bargaining game [75,76], sometimes referred to as a bargaining problem or bargaining game, two players could choose from an identical set of alternatives, however each alternative has varying payoffs for the players. Typically, some alternatives have better payoff for one player, while other alternatives have better payoff for the other player. If both players choose the same alternative, then each get the payoff corresponding to that alternative. If they choose differing alternatives, then there is no agreement, and they each get a fixed payoff which corresponds to the cost of non-agreement, and therefore typically negative. Thus, there is incentive to choose an alternative which may not necessarily be the best for a player. If there is perfect information, that is, the full set of alternatives and payoffs is known to both players, then there is an equilibrium solution to the Nash bargaining game [77].

2.2.12. Subgames

A subgame is a subset of a sequential game such that at its (the subgame's) beginning, every player knows all the actions of the players that preceded it [14]. That is, a subgame is a section of the game tree of a sequential game where the first decision node that belongs to this section has perfect information.

2.2.13. Subgame perfect Nash equilibrium

In a sequential game, a subgame perfect Nash equilibrium is a set of strategies representing each player such that they constitute a Nash equilibrium for every subgame of that sequential game [14]. Thus, when a game tree of the sequential game is considered, if a set of strategies could be identified so that they represent a Nash equilibrium for every branch of game tree originating from a node which represents a point where every player knows all preceding actions of all players, such a set of strategies represent a subgame perfect Nash equilibrium for that sequential game. For example, when two players are bargaining, they are in a subgame perfect Nash equilibrium if they are presently employing a set of strategies, which will represent a Nash equilibrium between them at any future stage of the bargaining process given that they are aware of the full history of the bargaining process up to that point.

2.2.14. Behavioural Game Theory

Behavioural Game Theory combines classical game theory with experimental economics and experimental psychiology, and in doing so, relaxes many simplifying assumptions made in classical game theory which are unrealistic. It deviations from several simplifying assumptions made in classical game theory such as perfect rationality[48,78], the independence axiom, and the non-consideration of altruism or fairness as motivators of human decision making[79,80]. We will show in this review that the approaches related to behavioral game theory are crucial in modelling military scenarios, such as in signaling games.

2.2.15. Evolutionary game theory

Evolutionary game theory is an outcome of the adoptation of game theory into the field of evolutionary biology [81]. Some of the critical questions asked in evolutionary game theory include: which populations/strategies are stable? which strategies can 'invade' (become popular) in populations where other strategies are prevalent? How do players respond to other players receiving or perceived to be

receiving better payoffs in an iterated game setting? etc. Evolutionary games are often modelled as iterative games where a population of players play the same game iteratively in a well-mixed or a spatially distributed environment [62,82].

A strategy can be identified as an evolutionary stable strategy (ESS) if, when prevalent, it has the potential to prevent any mutant strategy from percolating its environment [62]. Alternatively, an ESS is the strategy which, if adopted by a population in a given environment, cannot be invaded by any alternative strategy. Hence, there is no benefit for a player to switch from an ESS to another strategy. Therefore, essentially, an ESS ensures an extended Nash equilibrium. For a strategy S_1 to be ESS against another 'invading' strategy S_2, one of the two conditions mentioned below needs to be met, in terms of expected payoff E.

1. $E(S_1, S_1) > E(S_2, S_1)$: By unilaterally changing strategy to S_2, the player will lose out against another player who sticks with the ESS S_1.
2. $E(S_1, S_1) = E(S_2, S_1)$ & $E(S_1, S_2) > E(S_2, S_2)$: a player, by converting to S_2, neither gains nor loses against another player who sticks with the ESS S_1, but playing against a player who has already 'converted' to S_2, a player is better off playing the ESS S_1.

If either of these conditions are met, the new strategy S_2 is incapable of invading the existing strategy S_1, and thus, S_1 is an ESS against S_2. Evolutionary games are typically modelled as iterative games, whereby players in a population play the same game iteratively.

2.2.16. Signaling Games

A signaling game [83] is an incomplete information game where one player has perfect information and another does not. The player with perfect information (the Sender S) relays messages to the other player (the Receiver R) through signals, and the other player will act on those signals after inferring the information hidden in the messages. The Sender S has several potential *types*, of which the exact type t in the game is unknown to the Receiver R. t determines the payoff for S. R has only one type, and their payoff is known to both players.

The game is divided into the sending stage and the acting stage. S will send one of $M = \{m_1, m_2, m_3, ..., m_j\}$ messages. R will receive that message and respond with an action from a set $A = \{a_1, a_2, a_3, ..., a_k, \}$. The payoff that each player receives is determined by the combination of the Sender's type and message, as well as the action that the Receiver responds with. An example of the signaling game is the Beer-Quiche Game [83], in which Player B, the receiver, makes a choice of whether or not to duel Player A. Player A is either surly or wimpy, and Player B would only like to duel the latter. Player A chooses to have either a Beer or a Quiche for breakfast. While they prefer a quiche, a quiche signals information from a stereotype that quiche eaters are wimpy. Player B must analyse how each decision, duel or not duel, will give them payoff depending on which breakfast Player A chooses.

With this background, we now review the available literature which deals with the applications of game theory in defence science and technology.

3. Use of game theory in defence science and technology applications

As mentioned earlier, the primary parameters that influence the payoff matrix in games modelling defence scenarios are the value of targets, the value of resources and the priority of objectives. Other than this, the games used in defence applications can vary greatly, as we will see below. For this reason, this section is structured based on the domain (type of warfare) each paper covers Where a paper covers more than one domain, it is included in the most relevant subsection / domain. We however analyse in detail the type of games used, the way the payoff functions were structured, the available strategies and equilibria etc for each paper.

3.1. Papers dealing with Land Warfare

In Land warfare related applications of game theory, most studies focus on defensive warfare, whereby the military makes decisions on how to best allocate their ground defences to multiple threats. Some studies also focus on historical land-based conflicts and provide game theoretical analysis in hindsight, revealing how some decisions made from intuition in historical conflicts had a rational and mathematical justification. Land warfare can result in very heavy casualties, so understanding how to best minimise human losses is a key component (though not the only objective) of land warfare. Quite often, prioritising military resources is also fundamental to success and often features prominently in strategic decisions. Furthermore, often in scenarios involving ground warfare, it is important to assess the knowledge about opponents , their possible tactics, or terrain: it may become necessary to combat airborne forces being inserted at certain places, or it may be needed to traverse uncertain territory. In each of these situations, understanding where a force has imperfect information will help that force to make rational decisions.

Bier et al. [84] design a game to best assign defensive resources to a set of locations / resources that need to be protected. The attacker must then decide how they choose to split their force to attack the different targets. The game is modelled as a two player game of normal form. Payoff in this game is absolute, and an attack on a location i is either a success or a failure, where the attacker gains a_i and the defender loses d_i. Since orders for an attack are confirmed ahead of an attack, attackers must use a set of pure strategies. The game can be played both simultaneously or sequentially. That is, the game can be played depending on whether or not the attacker knows how the defender has assigned their resources before making their decision. This leads to the ideal strategy being to leave some targets undefended, centralising defenses in key areas by leaving some areas vulnerable.

There are several studies which analyse historical conflicts, which occurred predominantly on land, using a game theoretic prism. For example, Cotton and Liu's [85] describe two ancient Chinese military legends and model them as signaling games. In both games, legendary military leaders are faced with formidable opposing armies with much greater numbers and strength than their own, but instead of retreating, they prepare to engage, acting as if they are setting up for an ambush. Their opponents with imperfect information are left only with the messages they can infer from their opponents' actions; spooked by the perceived confidence and the reputation that these generals carried, the opposing armies, though in actuality are of superior strength, choose not to engage. Through a brave and ingenious bluff, both generals achieve an equilibrium solution in their favour by standing their ground. They do this by creating deception without direct communication, which follows the template of the aforementioned Beer-Quiche signaling game.

The first game described by Cotton and Liu is the "100 Horsemen" game. They describe a piece of history where a hundred Han horsemen travelling alone encounter a large Xiongnu force numbering in the thousands. Their available strategies are to retreat or engage. If they retreat, and the enemy engage, they will very likely be run down and defeated; and if they engage and the enemy also engage, they will be eliminated in battle. The best outcome for them is to somehow force an enemy retreat. The enemy is uncertain if the horsemen are travelling with a greater army. They see the horsemen move to engage, and decide not to take the risk, and retreat. The situation is translated into a two player game, with two strategies. It is represented in Figure 1 below:

In Figure 1 :

LG represents the decision point for the General Li Guang, of the Han forces.

GenX represents the decision point for the opposing Xiongnu force.

Payoffs are listed as (LG, GenX)

$\lambda \in (0, 1)$ represents the General's ability, as either strong or weak

α and β represents the proportion of Han horsemen killed in a retreat

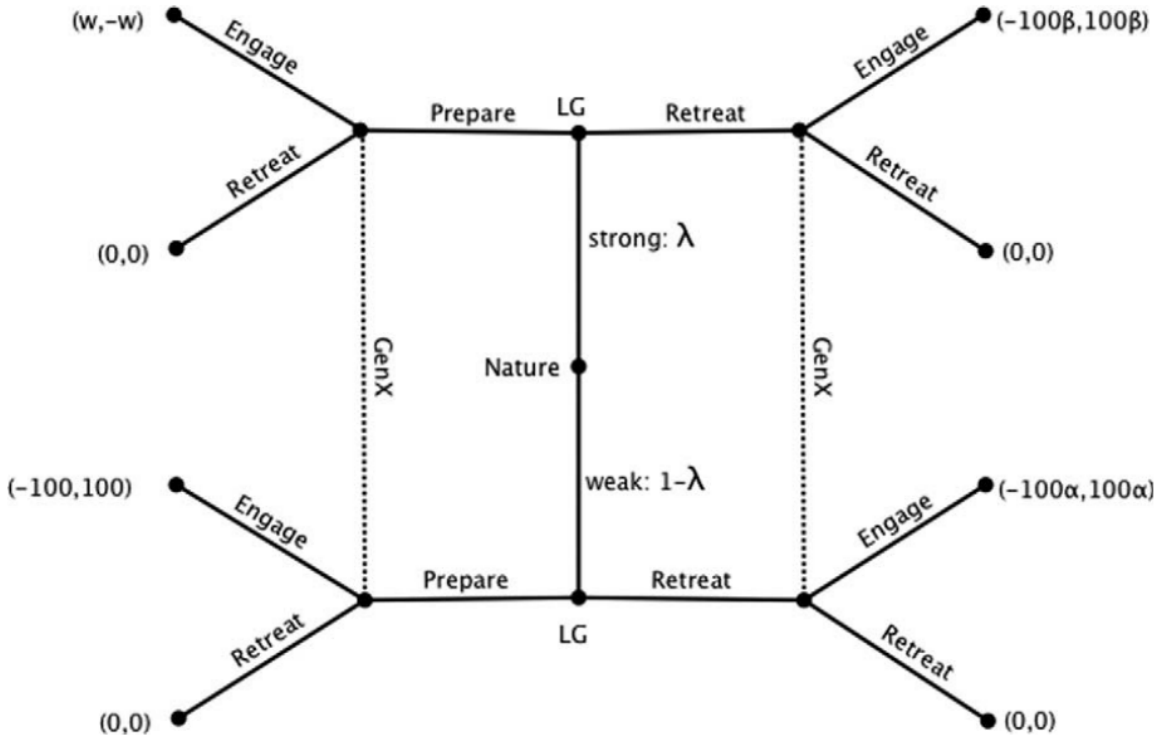


Figure 1. The 100 Horsemen signaling game [85]

w is a positive parameter

The second game is very similar to the first. In this game, a small city is guarded by the formidable General Zhuge Liang. He learns that a great hostile army is approaching the city. He is faced with two options. He could run, after which he would secede the city and likely be chased down by the approaching army, or he could stay and defend the city. If he chose the latter, and the army were to engage, he would likely lose his life, his army and the city. Faced with this dilemma, he orders his men to hide out of sight, so that the city appears empty from the outside. He climbs to the top of the foremost tower of the city and plays music. The opposing general, aware of General Liang's experience and prowess, suspects that the General has taken this unassuming position in the tower in the empty city to ambush his army, and they move away from the city to avoid being ambushed. General Liang sent effectively two signals here. The first was his reputation, a signal encompassing his strategic and military strength. The second was his choice to stay and defend the city. With these two pieces of information, and nothing else about the whereabouts or magnitude of General Liang's army, the opposing army chooses the safe option of zero loss and leaves. This piece of history is modelled as another two player signaling game, shown in Figure 2 below:

In Figure 2 :

ZL represents the decision point for the General Zhuge Liang

Payoffs are listed as (ZL, Opposing Army)

$\lambda \in (0, 1)$ represents the General's ability, as either strong or weak

c represents the value of the city

w represents the gains if ZL's army matched the opposing army

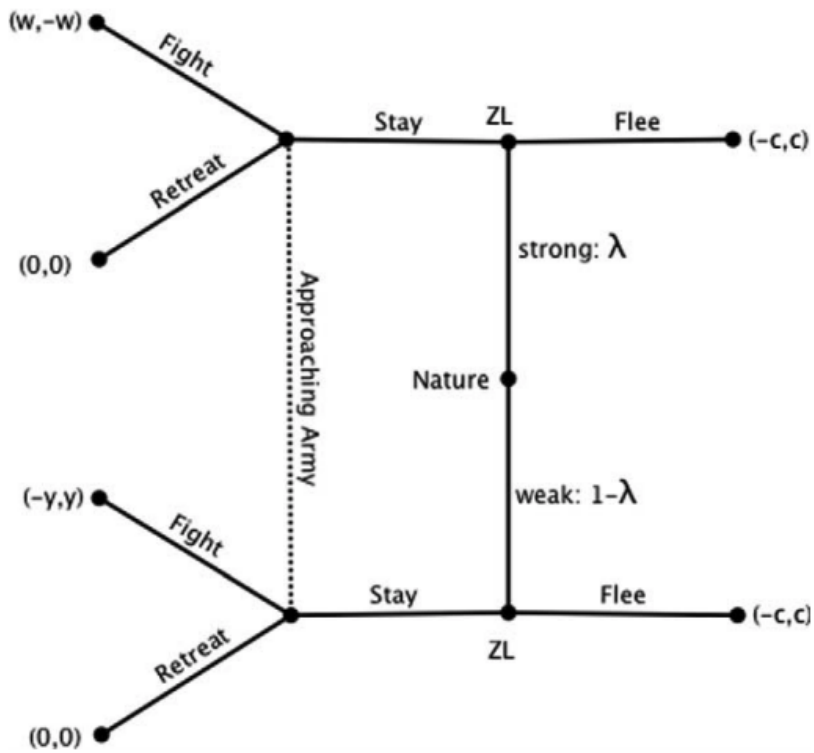


Figure 2. The Empty City signaling game [85]

y represents the losses if ZL's army is weaker than the opposing army, and $y > c$ since it encompasses losing the city

Both pieces of history represent distinguished military decision making in the face of near certain defeat, and are in fact examples of Generals with a strong understanding of the nuances of signals and rational decision making in strategic interactions forcing a favourable outcome to themselves.

The next paper we review is Gries et al [86] which is a comprehensive investigation into the utility of game theory principles in guerilla/destabilisation warfare. The significant factors they model are: destabilisation insurgents often attack randomly, creating a continuous threat that must have a continuous mitigation and detection strategy; duration of a war is important to consider, and will change the value that is assigned to targets and assets; time preferences play a critical role in setting priorities, as always judgements of value determine strategic decisions which in turn determine success or failure. The game model they propose involves both a sequential non cooperative game and a simultaneous non cooperative game, in each of which the two players are the guerilla force and the government. For these conflicts, the economic and social impacts are much more significant than military losses and gains, and therefore play a much more significant role in calculating the value of outcomes.

The game specifically looks at moments when each side looks to try and find peace or conflict with the other. At each of these moments, the government force must consider the financial cost of each option, while the rebels must on the order of priority of the engagements, and what portion of their fighting force they will make available for each engagement. Figure 3 demonstrates an example of the decision tree to emerge from these moments in destabilisation warfare, where G represents the Government decisions and R represents the Rebel decisions.

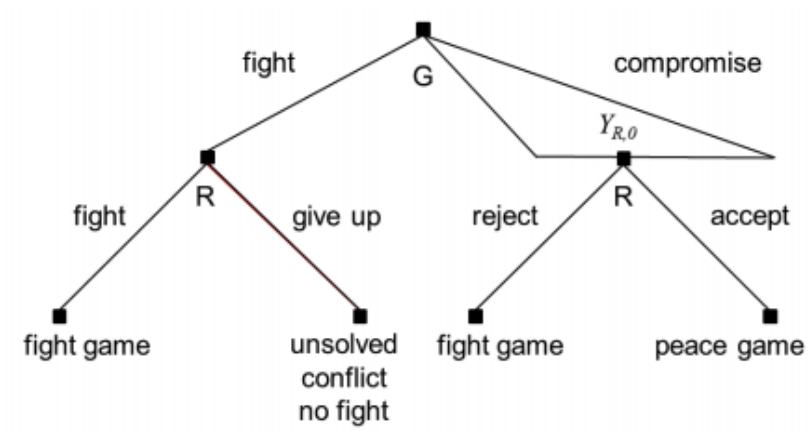


Figure 3. Destabilisation Warfare game [86]

Krisnamurthy et al. [87] investigate a game-theoretic control of dynamic behaviour of an unattended ground sensor network (UGSN) in order to acquire information about intruders. Each sensor in this network is capable of receiving measurements, with specified accuracy, of the range and bearing of nearby targets which they then transmit to a local hub for data fusion. In this framework, while more sensor measurements and larger volumes of transmission of measurements may lead to better target awareness, this also results in the undesirable effect—greater consumption of limited battery power. Hence, the goal to which game-theory is applied is to optimally trade-off target awareness, data transmission and energy consumption using a two-time scale, hierarchical approach.

The authors demonstrate that the sensor activation and transmission scheduling problem can be decomposed into two coupled decentralized algorithms. In particular, the sensors are viewed as players in a non-cooperative game and an adaptive learning strategy is proposed to activate the sensors according to their proximity to targets of interest. This turns out to be a correlated equilibrium solution of this non-cooperative game. Next, the transmission scheduling problem, in which each sensor has to decide at each time instant, whether to transmit data and waste battery power, or to wait and increase delay, is formulated as a Markov Decision process with a penalty terminal cost. The main result of this formulation is to show that the optimal transmission policy has a threshold structure which is then proved using the concept of supermodularity.

3.2. Papers dealing with Naval Warfare

Naval warfare predates aerial warfare and has been prevalent for a longer duration in human history. In the 18th and 19th centuries, the powerful nations of the time built warships with cannons positioned along their sides. It meant that ships could attack typically only to their sides. When sailing as an armada, the standard approach was to form a 'line of battle' i.e. a column of allied naval ships sailing in a direction such that their sides would face the enemy, also positioned in a line. The two parallel opposing fleets could blast one another with large number of cannons. The 'line of battle' strategy is considered to be a Nash-equilibrium because neither fleet would gain from performing raking (a tactic of the era, whereby an attacking ship would attempt to sail across its adversary's stern, concentrating cannon fire there while the enemy could only respond minimally due to having less cannon defences in the stern. The attacking ship would damage both the stern and some of the broadsides of its adversary). According to [88] raking was not preferred in a fleet as this would mean having to first sail ahead of its enemy and then turn towards

it—a challenging task when the ships' speeds were roughly equal and maneuvering was difficult. As neither fleet would gain from turning towards the enemy and neither would get ahead, Levine concludes that this strategy—forming a line of battle and sailing parallel to the other fleet—was each fleet's best response, and thus represented a Nash equilibrium.

Levine goes on to mention battles in which English fleets deviated from the above strategy and sailed orthogonally towards a French and Franco-Spanish fleet. In the first battle Levine mentions, it was likely unplanned. In the second—the 1805 Battle of Trafalgar—it was by careful design: the English fleet divided itself into two columns, each of which sailed orthogonally to the Franco-Spanish line, raking fire for about 45 minutes before crashing through it and beginning a general melee. The English would go on to isolate the middle of the Franco-Spanish fleet to score a decisive victory. Levine considers both battles to be counterexamples to his thesis. However, in the Battle of Trafalgar, its is possible that the English strategy was a best response to the likely Franco-Spanish strategy of forming an orthodox line of battle. The English admiral, Lord Nelson, desired to keep the Franco-Spanish fleet from escaping—which they could if both fleets formed parallel lines of battle—thus reducing the reward he would get for forming his own fleet into a line of battle. Moreover, he may have estimated that the poor gunnery of the French and Spanish ships would lessen the effect of the raking fire, thus reducing the negative reward he would get for directly charging the Franco-Spanish fleet. In his eyes, this may have made the unorthodox option a better response to the likely Franco-Spanish strategy than the orthodox line of battle. While Levine did not explicitly attribute these strategies in naval battles of the era to game theory, it is quite possible that the adopted strategies had a game-theoretic basis.

Maskery et al. [89] study the problem of deploying counter-measures against anti-ship missiles using a network enabled operations (NEOPS) framework, where multiple ships communicate and coordinate to defend against a missile threat. Here, the missile threats are modelled as a discrete Markov process and they appear at random positions within a fixed physical space and move towards the ships obeying some known target dynamics and guidance laws. The ships which are equipped with counter-measures (CM) such as decoys and electromagnetic jamming signals, are modelled as the players of a transient stochastic game, where the actions of the individual players include the use of CM to maximize their own safety while cooperating with other players which are essentially aiming to achieve the same objective. The optimal strategy of this game-theoretic problem is a correlated equilibrium strategy, and is shown to be achieved via an optimization problem with bilinear constraints. This is in contrast to the Nash equilibrium solution proposed in [90] to a related problem but one without player coordination. A noteworthy contribution of this paper is that it also quantifies the amount of communication necessary in order to implement the NEOPS equilibrium strategy. This paper highlights the utility of game-theoretic methods in analysing optimal strategies in network enabled systems which are critical in modern warfare.

In [90] Maskery et al. consider the problem of network centric force protection of a task group against anti-ship missiles. The decision makers in this model are the ships equipped with hard-kill/soft-kill weapons (counter-measures) and these ships are also considered the players in the formulation of this problem as a game-theoretic setting. The platforms must make critical decisions independently on the optimal deployment of the counter-measures while they simultaneously work towards a common goal of protecting the members of the task group. Essentially this is a decentralised missile deflection problem in a naval setting which is formulated as a transient stochastic game for which the ships may compute a joint counter-measure policy that is in Nash equilibrium. Here the ships play a game with each other instead of with a missile. This approach naturally lends itself to decentralised solutions which may be implemented when full communication is not feasible. Moreover, this formulation leads to an interpretation of the problem as a stochastic shortest past game for which Nash Equilibrium solutions are known to exist.

Bachmann et al. [91] analyse the interaction between a radar and jammer using a noncooperative two-player, zero-sum game. In their approach, the radar and jammer are considered "players" with opposing goals: the radar tries to maximize the probability of detection of the target while the jammer attempts to minimize its detection by the radar by jamming it. Bachmann et al. [91] assume a Swerling Type II target in the presence of Rayleigh distributed clutter, for which certain utility functions are described for cell-averaging (CA) and order-statistic (OS) CFAR processors in different cases of jamming. This game-theoretic formulation is solved by optimizing these utility functions subject to constraints in the control variables (strategies), which for the jammer are jammer power and the spatial extent of jamming while for the radar the available strategies include the threshold parameter and reference window size. The resulting matrix-form games are solved for optimal strategies of both radar and jammer from which they identify conditions under which the radar and jammer are effective in achieving their individual goals.

3.3. Papers dealing with Aerial Warfare

Aerial combat is often a normal-form game where decisions about utilised resources are made before the engagement, based on assumptions and knowledge about the strength of different elements of the arsenal. For example, Suppression of Enemy Air Defense vehicles (SEADs) are effective against ground-to-air defenses and Surface to Air Missiles (SAMs), but will not be useful against fighter aircraft. Therefore, when military personnel decide which resources to use in an engagement, they need to weigh how valuable each of their resources is, as well as how important the objective is to both sides of the conflict. If the attacking force values a target much more than it is actually worth, then their increased resource expenditure may be detrimental to their military campaign as a whole. With humans operating the aerial weaponry usually, their respective abilities and skill sets need to be considered, and the likelihood of them executing their mission.

There is limited literature on aerial combat modelled with game theory. Hamilton [92] writes a comprehensive guide to applications of game theory on multiple Aerial warfare situations. Hamilton suggests using game theory to devise strategies not only based on one's own military options, but also expectations around enemy actions as well. Game theory accounts for different interactions with the enemy, rather than simply considering which side had superior maximum-effort power. Nowadays, many military forces are able to adapt to instantly changing situations and adjust their actions based on those new circumstances. As such, Hamilton suggests first determining all of the tactical options available to each side. As stated earlier, one of the most fundamental elements of using game theory for the military is understanding exactly how much value each asset holds - and detailing the inventory and strategic possibilities of both sides will best clarify all strategic options. For each option, Hamilton suggests assigning a numerical value - a Measure of Effectiveness (MoE). Decisions about MoEs are important because being accurate with MoEs will underpin the choices that are made strategically. Incorrect MoEs can lead to incorrect strategic decisions, and perhaps also resulting in poor understanding of why the decision was wrong. An example of this (although not in the aerial warfare context) was the Vietnam War, where the early US strategy was maximising the neutralisation of Viet Cong soldiers. Since the Northern Vietnamese leadership did not place great emphasis on their infantry, the US strategy ultimately led to a loss in the war. Next, Hamilton suggests calculating the combined value for all possible interactions between strategies of both sides of the conflict. This will generate a matrix of payoffs, from which it is possible to derive the optimum or dominant strategy for each player, and then an equilibrium solution. Thus, ahead of any engagement a military leader may partake in, they have a well-formed idea of the expected result of the game. A caveat that Hamilton adds to these guidelines is to consider the length of a military campaign as a whole. The values that can be assigned to a resource for one battle or strike attack

may be small if they are low cost or numerous asset, but depending on how many of these skirmishes happen over the course of a campaign, those resources may be come pivotal over the course of a war.

To illustrate these points, Hamilton applies them to a standard aerial warfare game of SEADs and time critical targets. In this combat, the 'Blue' side is trying to eliminate some ground based targets. To do this, they use SEADs. In response, the 'Red' side will fire SAMs, which SEADs struggle to avoid. However, expecting this response, the Blue side also has Strike aircraft which can defend the SEADs and counteract the SAMs, but are unable to attack the targets. The questions for the Blue team are: what is the value of the target and what ratio of SEADs and Strike aircraft should be deployed for the targets? Likewise, for the Red team: how valuable is the target and how many, if any SAMs should be fired? Hamilton contends that the optimal Red strategy is to fire only for a fraction of engagement which is equal to:

$$\frac{\text{Value of Target}}{\text{Value of Target} + \text{Value of SEAD } \times Pk_s + \text{Value of SAM } \times Pk_A} \tag{2}$$

and the optimal Blue Strategy is to assign a fraction of the planes as SEADs which is equal to:

$$\frac{\text{Value of SAM } \times Pk_A}{\text{Value of SAM } \times Pk_A + \text{Value of SEAD } \times Pk_s + \text{Value of Target}} \tag{3}$$

where

Pk_s is the probability of the SAMS destroying the SEADs

Pk_A is the probability of the Strike aircraft destroying the SAMs

This formulation gives a concise prediction of the likely outcome of an engagement given every possible assignment of aircraft and missile launches. It must be noted that it is incredibly difficult in practice to accurately quantify the numerical value of different targets and resources.

Deligiannis et al. [93] consider a competitive power allocation problem in a MIMO radar network in the presence of multiple jammers. The main objective of the radar network is to minimize the total power emitted by the radar while achieving a specific detection criterion for each of the targets. In this problem, the radars are confronted by intelligent jammers that can observe the radar transmitted power and thereby decide its jamming power to maximize interference to the radar. Here they treat this power allocation problem as a non-cooperative game where the players are the central radar controller and the jammers, and solve this using convex optimization techniques. Moreover, they provide a proof for the existence and uniqueness of the Nash equilibrium stable point, where no player can further profit by changing its power allocation.

Garcia et al. [94] investigate the problem of defending a maritime coastline against two enemy aircraft whose main objective is to invade the territory controlled by the defending aircraft. The defender, on the other hand, attempts to prevent this by trying to intercept both enemy aircraft in succession as far as possible from the border. This is a typical pursuit-evasion scenario and is representative of many important problems in robotics, control and defence. In this paper, Garcia et al. formulate this problem as a zero-sum differential game, where the defender/pursuer tries to successively capture the two attackers/evaders as far as possible from the defended coastline while the attackers cooperate and minimize their combined distance from the border before they are captured. They then find the optimal strategies for the attackers and the defender in this one-defender two-attacker pursuit-evasion game by solving a set of nonlinear equations. The cooperative strategy discussed in this paper provides an important coordination approach for less capable (perhaps slower) agents when they are tasked to carry out a mission.

He et al [95] consider the radar counter-measure problem in a multistatic radar network, where a game-theoretic formulation of joint power allocation and beamforming is studied in in the presence of a smart jammer. The goal of each radar in this network is to meet the expected detection performance of the target while minimizing its total transmit power and mitigate the potential interferences. On the other hand, the goal of the jammer is to adjust its own transmit power to interfere the radar so as to protect the target from detection. First, they study the power allocation game with strategy sets of each player (radar and jammer) consisting of their respective transmit powers. For this problem, they proceed to solve the corresponding optimization problems to work out the best response function for the radar and the jammer and show the existence and uniqueness of the Nash equilibria. Next, they consider the joint power allocation and beamformer design problem in the presence of jammers again as a non-cooperative game and propose a power allocation and beamforming algorithm which is shown to converge to its Nash equilibrium point.

McEneaney et al. [96] investigate the command and control problem for unmanned air vehicles (UCAVs) against ground-based targets and defensive units such as surface-to-air missile (SAM) systems. The motivation for this work arises from the requirement for operations planning and real-time scheduling in an unmanned air operations scenario. The problem is modelled as a stochastic game between blue players (UCAVs) and red players that comprise the SAMs and ground based targets. The game objective may vary, an example of which for a blue player is to destroy a strategic target while minimizing damage to itself. The red players, on the other hand, attempt to inflict maximum damage on the UCAV while protecting themselves from attack by the UCAVs.

The control strategies for the UCAVs consist of a set of discrete variables that correspond to the specific target or SAM to attack while that for the SAMs are to switch their radar "on" or "off". Note that when the radar is "on", the probability of the SAM causing damage to the blue players increases as does the probability that the blue players inflict damage on the SAM. The solution to this stochastic game is obtained via dynamic programming and illustrated on some numerical examples. A main contribution of this work is the analysis of a risk-sensitive control based approach for stochastic games under imperfect knowledge. In particular, this approach not only handles noisy observations due to random noise, but also deals with cases that include an adversarial component in the observations.

Wei et al. [97] have developed a mission decision making system for multiple uninhabited combat aerial vehicles (UCAVs) working together. The UCAVs weapons are air-to-air missiles. In the paper a red-UCAV team consisting of an unmanned fighter-bomber flanked by two UCAVs attempts to strike a blue-team ground target. The blue-team has its own set of UCAVs that are directed to defeat the red-team. The success of a given missile against its chosen threat is determined by the distance between the attacker and threats, their relative speed, and relative angles. The blue team versus red-team scenario is represented as a simultaneous normal form game with the strategies for the team corresponding with allocations of blue team entities against red-team entities and vice versa. In the paper the payoff for red or blue team is based upon considering the effectiveness and ineffectiveness of a given allocation, which is turn is dependant upon the relative geometry between the opposing team allocation groupings. Dempster-Shafer (D-S) theory is applied where the D-S combinatorial formula is harnessed to formulate the payoff. These payoffs, calculated for each strategy, for each team is then placed into bi-matrices i.e., one for each team and solved using a linear programming optimisation approach. If an optimal Nash equilibrium point is not present, mixed strategy approach is applied and solved for. The authors have developed some mission scenarios with differing geometries and illustrated the use of their game-theoretical allocation strategy. They use annotated diagrams of entity geometry containing red and blue teams in proximity to

one another to justify that the allocation strategy determined by their payoff formulation is satisfactory.

Ma et al. [98] have developed a game-theoretic approach to generate a cooperative occupancy decision-making method for multiple unmanned aerial vehicle (UAV) teams engaged against each other in a beyond-visual-range (BVR) air combat confrontation. BVR combat is targeted because of developments in missile technology enabling long-range engagements. In the paper, the team on each side first decides the occupancy positions (cubes in Cartesian space) of its UAV entities followed by selecting targets for each UAV team member to engage. The goal is for each side to obtain the greatest predominance while experiencing the smallest possible threat condition. A zero-sum simultaneous bi-matrix game is applied to express the problem. For a given occupancy of a UAV, height and distance predominance formulas that factor in range and weapon minimum/maximum performance criteria are used to generate payoff values for the utility functions. Since, the scale of the game leads to a explosion in size (and thus strategy) as the number of occupancy cubes and UAVs for each team is increased, the authors have chosen to augment the Double Oracle (DO) algorithm, that was designed to solve large scale zero sum game problems in earlier works by combining it with a neighbourhood search (DO-NS) algorithm. Through simulations, the authors illustrate that the results show the DO-NS algorithm outperforming the DO algorithm in terms of computational time and solution quality.

Evers et al. [99] have developed the design for a Cooperative Ballistic Missile Defence Game (CBMDG). This game has been designed to support strategic negotiations between a threatened nation and a possible coalition of nations that can be integrated together to form a layered BMD to protect the threatened nation. Through the game, the assignment of ballistic missile interceptors to the coalition nations is determined which minimizes the expected number of interceptors required to achieve a desired defence level in case of an attack from an enemy. The paper shows that the number of interceptor required by a coalition is smaller compared to utilising interceptors solely within the threatened nation to engage the BM threat, hence motivating it to seek out a coalition based layered defence. For the negotiation, the cost savings that threatened nation gains due to having to spend less on interceptors by leveraging cooperation via a coalition corresponds with the bargaining power that the nations in the coalition can demand as compensation. Using the interceptor savings game formulation devised in the paper, the savings introduced through each nation participating in the coalition can be determined and the fair compensation can be allocated to each aiding participating nation.

Garcia et al. [100] consider an air combat scenario where a target aircraft that is engaged by an attacking missiles utilises a defending missiles as a countermeasure to defend itself as it attempts to escape the attacker by maximising my the distance between itself and the attacker when the defender reaches as close at it can to the attacking missile. The game is referred to as an active target defense differential game (ATDDG). In the paper, the authors extend previous works performed on this three party problem to develop a closed-form analytical solution for the ATDDG where the Defender missile is able to defeat the attacker if it enters within a with a capture circle with specified radius $rc > 0$. Additionally, the closed-form optimal state feedback solution generated in the paper is supposed to work in spite of the attacker employing an unknown guidance law rather than assuming it is Propotional Navigation (PN) or pursuit (P). Finally, the authors provide the set of initial conditions for the target aircraft where its survival is guaranteed if the target-defender team plays optimally depsite the unknown guidance law employed by the attacking missile.

Han et al. [101] present an Integrated Air and Missile Defence (IADS) problem where surface-to-air (SAM) batteries equipped with interceptor missiles (IM) engage the attacker missiles (AM) targeted at damaging cities being protected. The problem is cast as a simplified two-party zero sum (equally valued

targets) game with perfect information and has three stages. By perfect information, it is implied that both players know what has happened in the previous stages of the game The three stages correspond with defender setting up their allocation of SAMs to cities, followed by attacker allocating their missile salvo against cities and finally the defender in response allocating interceptor missiles to counter attacker missile salvo. The simplifying assumptions made in this problem are that there is only one SAM allocated near a city and only one installed per site. Additionally, no more than one interceptor is launched against each attacking missile. Additionally, only one IM can be allocated to one DM, each each SAM has the same number and type of IMs, AMs are identical and are fired in a single salvo. It is attempted to be solve the tri-level game using an extensive form game tree, $\alpha - \beta$ pruning and using the Double Oracle (DO) algorithm for a 6 city network that needs to be protected. Only DO is a heuristic based approach and is not guaranteed to find the SPNE. The efficiency with which sub-game Nash equilibrium is reached by each choice of algorithm is studied. For the game-tree approach, the conclusion made is that size of the strategy space is determined to increase to an intractable size because of the combinatorial nature of the problem. When applying $\alpha - \beta$ pruning, the author determines that the when scaling to change number to increase SAM batteries, AMs and IMs, it does not scale well in terms of computation time compared to DO. However DO does fail to find the SPNE in a small number of instances. The author preferences DO despite its lack of guarantee to reach the SPNE since it is shown to not violate monotonicity (increase in payoff) and the solution quality trends (non exponential increase in computational time) even when increasing the size of the problem from 6 cities to 55 cities.

The work of Başpınar, Barış et al. [102] focuses on modeling of air-to-air combat between two unmanned aerial vehicles (UAVs) using an optimization-based control and game-theoretic approach. In this work, the ability to provide a controller despite the presence of complex non-linear dynamics for a UAV is achieved through avoiding integration of the differential equations involved. It is possible to do this because the differential flatness theory can be successfully harnessed to describe the movements of the UAV in the horizontal and vertical planes. In particular, vehicle motion is expressed in terms of specific variables and their derivatives through parameterized curves. Then, any trajectory planning, for moving from one waypoint to another can be solved by determining the smooth curves satisfying defined conditions in flat output space. Following determination, all variables involved to describe the smooth curve can be reverted to the original state/input space. The impact is a speed-up in the solving of any trajectory optimization through reducing the number of variables required. Game theory is then harnessed where the aerial combat between the two UAVs is modelled as a zero-sum game using a minimax approach. That is the each party tries to maximise its payoff when the opponent plays its best strategy. Here, The objective is for each UAV is to get directly behind the other party and within a range threshold in order to satisfy onboard weapon effective range constraints. In the paper, the authors provide cost functions associated with degree of being in tail-chase to the target based on aspect and bearing angles as well as the cost functions associated with generating a maximum score when the opponent is within some threshold of the optimum shooting range. The cost functions are multiplied together to create the total cost. The cost functions are put into a receding horizon control scheme where the trajectory planning determined through selection of the controls is performed for a given look-ahead time period where both players are utilising opposite strategies. Each player considers its opponents reachable sets within the horizon and uses this to select its choice of controls to maximise its payoff. This process is repeated every few control steps. For selecting the control actions, that authors mention the use the full set of control inputs within the performance envelope rather than subset (e.g. turn, maintain hading, roll left at particular angle, immelman, split S or spiral dive) unlike most other works and thus point to generating a more optimal solution for each player selecting their respective strategies. Two simulation scenarios are provided with the first being the case where neither UAV starts off in air-superiority position and then exercises the receding horizon cost function optimisation to get into tail-chase with its opponent within

optimum firing range. Neither succeed. The authors show the speed, load factor and bank-angle when applying the controls do not violate bounds during the flights and that feasible trajectories are generated. For the second simulation, the UAVs are initially in a tail-chase except not satisfying the within shooting range criterion. The opponent being chased manevouers to escape through applying the cost function while the chaser continues chasing. At the end of the engagement, within shooting range criteria are met and the target is directly in front but at the sub-optimal aspect, which leads to its escape. These scenarios are used to demonstrate the validity of the control strategy developed and thus provide the automatic selection of combat strategy for two unmanned aerial vehicles engaged in combat against one another.

Casbeer et al. [103], consider a scenario where an attacker missile pursuing an unmanned aerial vehicle target is engaged by two defending missiles launched from entities allied to the target which cooperate with the target. So, this scenario is an extension from a three party game where there is only one single defending missile engaging the attacker that cooperates with the target. The three party game is referred to as the Active Target Defence Differential Game (ATDDG). Besides computing the optimal strategy for the players in the extension to ATDDG, the paper is directed to determine the degree of reduction in vulnerability of the target when it uses two defenders rather than one. A constrained optimisation problem is formulated to setup the extension. It is shown that the target through having the choice to cooperate with either defender can more successfully escape the attacker. Additionally the presence of two defenders enables the attacker to be more easily intercepted. When the two defender missiles ae well positioned, both are able to intercept the attacker.

3.4. Papers dealing with Cyber Warfare

Keith et al. [104] consider a multi-domain (cyber combined with air-defence) defence security game problem. There are two players engaging each other in a zero sum extensive form game, a defender, representing an integrated air-defence system combined with cyber security protection and an attacker, capable of unleashing air-to-ground threats (missiles, bombs) as well cyber-attacks (against IADS network). Here, the payoff has been selected as the expected loss of life. The defender wants to minimise this while attacker wants to maximise it. The cyber security game problem to protect the IADS is nested within the physical security game problem. The actions of the players correspond with allocations to activate IADS/cyber security responses nodes corresponding with population centres for the defender and allocations to attack IADS/associated cyber-security nodes by the attacker. The realism of the game is increased through provisioning in imperfect information, which is where the defender and attacker are not fully aware of vulnerability of nodes (selected randomly by a nature model which picks them in each game). Additionally, the defender is only able to sense cyber attacks on nodes probabilistically which has the implication that its allocation of cyber defence teams to particular IADS is only effective probabilistically. For the attacker, it can also determine the effectiveness of its cyber-attacks following physically attacking a node. This work is directed to advances security game literature by introducing the integrated domain and multiple periods for agent actions as well enabling continuous mixed form strategies. Also, the author considers it the first work where Monte Carlo (MC), discounted and robust counterfactural regret minimisation based approaches have been compared in security game that are applied and evaluated following increasing scale of the problem. Initially, for a small-scale problem the Nash equilibrium (NE) in the form of a sequence-form linear program is determined for the defender. Then, the problem is gradually scaled to expand to number of population centres to be defended up to an upper limit. Then the author, applies the approximate counterfactual regret minimisation (CFR) algorithm to reduce computation time while preserving the optimality of the strategy as much as possible. When the

scale is further increased a discounted CFR is introduce which further reduces computation time. The parameter space of the problem and algorithm are explored to select best choice of tuning parameters and extract best performance per algorithm. The impact of limiting the rationality of the players, through introducing bounded rationality where the players do not necessarily make optimal robust best response moves. They can only manage approximate robust best response moves. A robust best response for a player is defined as the compromise between the completely conservative NE strategy and tje completely aggressive best response strategy. It means that the strategy of the players has an exploitability where the opponent can capitalise on non-best responses being played by them. With respect to a player, the capability of their strategy to capitalise over opponent is referred to as exploitation. Conversely, the vulnerability of their strategy with respect to an opponent is referred to as exploitability. When running all the different algorithms, the results show the Nash equilibrium solution being the safest strategy since the best moves are being played which are neither exploitable. The performance charts show that the robust linear program generates the highest mean utility and highest exploitation/exploitabiltiy ration while also introducing the maximum computational time. The Data biased CFR is seen to offer the best trade-ff by offering a high mean utility, a explotiation to exploitability ration in favour of exploitation while producing the lowest computational time.

3.5. Papers dealing with Space Warfare

In the domain of space warfare, human resources and risks are much less prevalent here, and instead the focus is more on network strength and interaction between independent autonomous agents, connected or otherwise. Ultimately, warfare in these aspects will operate at a pace and in dimensions far beyond human cognitive capacity. Since the rapidity and complexity of decisions within engagements will almost certainly outscale military personnel's understanding, game theory will take the place of decision makers as part of the overall software system, and imbue future technology to take human/ social factors into consideration when making calculations. With a greater focus on connectivity and networking, the key to success in these areas relies on effective communication channels and a unanimously shared goal across the system.

Zhong et al [105] set the ambitious goal of optimising bandwidth allocation and transmission power across a satellite network. They base their research on bargaining game theory, and have to achieve compromise across: interference constraints, Quality of Service requirements, channel conditions, and transmission and reception capabilities for satellites at every point in the network. Interference constraints and bandwidth limitation are the surplus that need to be negotiated in the bargaining game, with each satellite using different strategies to improve their utility/ share of resources. This quickly escalates in complexity, with the most important takeaway from the model being the mapping of a problem to the cooperative bargaining game framework.

Similarly, Qiao and Zhao [106] detail some key issues with the finite energy availability for the nodes in satellite networks. Their paper offers a solution through a game theoretical model of a routing algorithm, and use it to find an equilibrium solution to the uneven network flow. The model locates certain network hot spots which are reserving a lot of energy, and takes measures to evenly distribute the resources. This is another case of a bargaining/ cooperative game across multiple players in a network.

3.6. Papers dealing with Target Tracking

Gu et al. [107] study the problem of tracking a moving target using a sensor network comprising of sensors capable of providing some position-related target measurements. Each sensor node has a sensor to observe the target and a processor to estimate its state. While there is some communication available among sensors, this ability is limited in the sense that each sensor node can only communicate

with its neighbors. The problem is further compounded by the fact that the target is an intelligent agent capable of minimizing its detectability by the adversary and thereby has the potential to increase the tracking error of the tracking agent. Gu et al. [107] solve this problem within the framework of a zero-sum game, and by minimizing the tracking agent's estimation error, a robust minimax filter is developed. Moreover, in order to handle the limited communication capability of the sensor nodes, they propose a distributed version of this filter for which each node only requires information in the form of current measurement and estimated state from its immediate neighbors. They then demonstrate the performance of their algorithm on a simulated scenario with an intelligent target and show that while the standard Kalman filter errors diverge, the minimax filter which takes into account the adversary's noise, is able to significantly outperform the Kalman filter.

Qilong et al.[108] similarly address the issue of tracking an intelligent target, but they model a scenario where the tracking players are also in pursuit, and the focus is on protecting the target. Additionally, the target is able to fire a defensive missile at the attacker/ tracker. The attacker has line of sight of both the target and the defensive missile. The target's plan is to allow the tracker to slowly close the distance between itself and the target, all the while maneuvering to develop and understanding of how the attacker reacts. When the attacker is close to collision, the defensive missile is released. The target and the missile then communicate, use the knowledge of the attacker's movement patterns, and adhere to an optimal linear guidance law to destroy the attacker. This was modelled as a zero sum competitive game between the attacker, the target, and tjhe defensive missile. However, the paper also focuses on the cooperative game played between the target and the defensive missile, which is a non-zero sum game. For them, the payoff is calculated by minimised miss distance (which ideally equals zero - a collision with the attacker), as well as the control effort required to guide the defensive missile.

Faruqi [109] discusses the general problem of applying differential game theory to missile guidance. They state that missile trajectory follows Proportional Navigation (PN), a set of navigational systems using specific guidance laws for tracking targets. The performance of these systems is measured by a Linear System Quadratic Performance Index (LQPI). With respect to differential game theory, they model the missile guidance problem by representing the missile navigation and trajectory with a set of differential equations. The general form of this problem is:

$$\dot{\underline{x}}_{ij} = F\underline{x}_{ij}(t) + G\underline{u}_i(t) - G\underline{u}_j(t) \tag{4}$$

and

$$J(...) = \frac{1}{2}\underline{x}_{ij}^T(t_f)S\underline{x}_{ij}(t_f) + \frac{1}{2}\int_{t_0} t_f[\underline{x}_{ij}^T QG\underline{u}_i + \underline{u}_i^T R_i\underline{u}_i + \underline{u}_j^T R_j\underline{u}_j] \tag{5}$$

where
$\underline{x}_{ij}(t) = x_i(t) - x_j(t)$: is the relative state of player i w.r.t player j
$\underline{u}_i(t)$: is the player i input
$\underline{u}_j(t)$: is the player j input
F): is the state coefficient matrix
G: is the player input coefficient matrix
Q): is the Performance Index (PI) weightings matrix for current relative states
S): is the PI weightings matrix for final relative states
R_i, R_j: PI weightings matrices on inputs

Faruqi mainly focuses on two player and three player games, while the utility functions are modelled based on the relative distance vectors between missiles and targets. Faruqi shows that game theory can be effectively used in missile guidance tasks involving PNs in modern missiles.

Evers [110] on the other hand analyses defence against Theater Ballistic Missiles (TBMs) using game theory. The proliferation of ballistic missiles and nuclear technology has important consequences for military conflict, where the cost of failure can lead to the destruction of entire cities. It can be hard to pinpoint their launch, since they have a big range are very powerful, though their payload can vary considerably. In combating this threat, the defending nation does have the advantage that there is usually a long flight trajectory, typically split into three phases, during which the TBM can be intercepted. The Boost phase marks launch and the majority of the TBM ascent. The end of the boost phase is marked by the burn-out after which the TBM enters its midcourse phase. This phase is the longest phase of the flight, and affords the best opportunity for defenders to intercept the TBM. After the midcourse phase, the TBM enters its terminal phase from re-entry into the atmosphere. This is the last opportunity for defenders to intercept the missile. The flight path is illustrated in Figure 4 below.

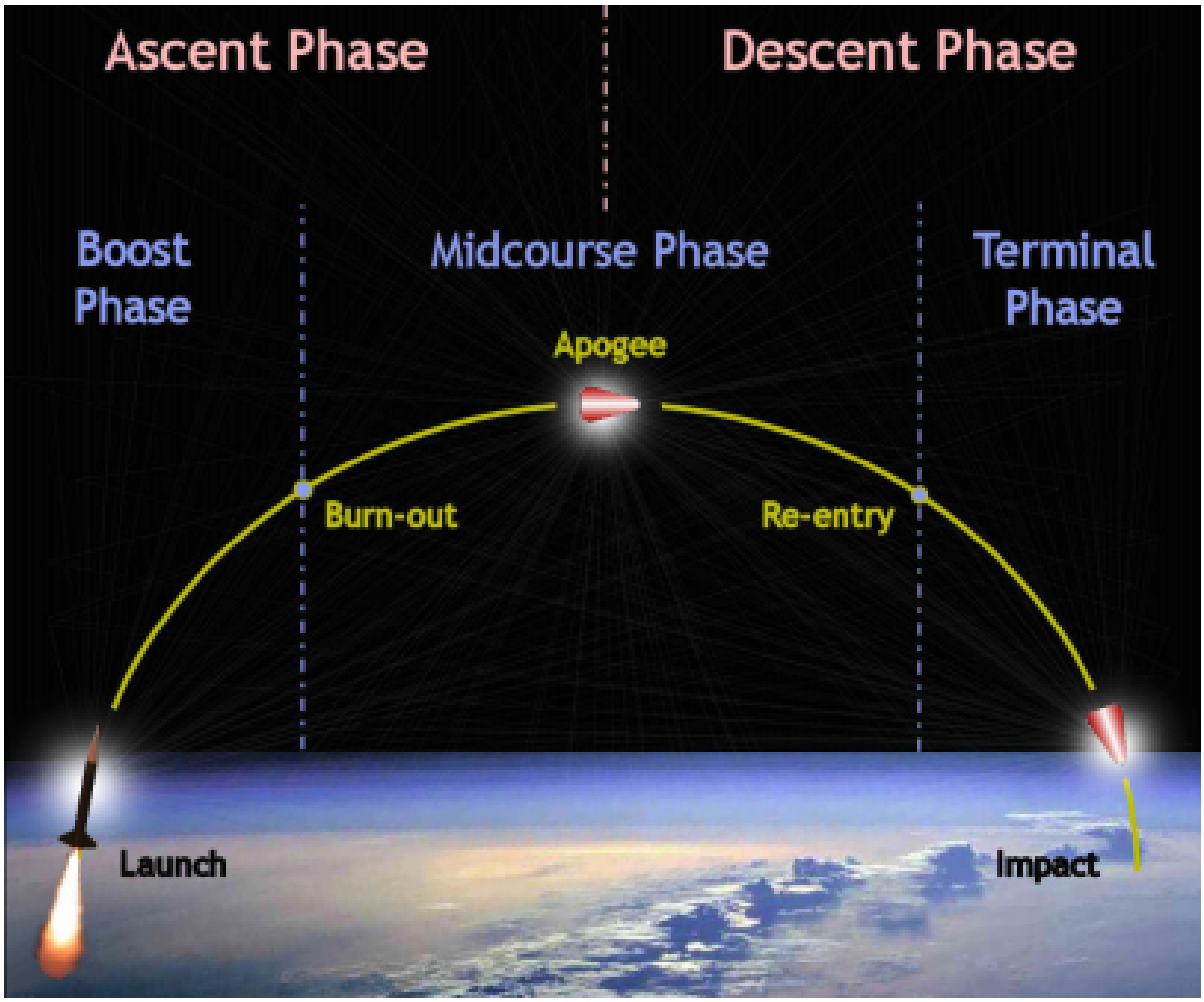


Figure 4. Flight path of Theater Ballistic Missiles [110]

The missile travels a long distance over a reasonably extended flight time. However, from its physical geographical location, a defending military force or nation can only apply its resources to defend against it in the termination phase of flight, where the risk is much higher and the cost of failure is at its greatest. For this reason, Evers proposes a cooperative strategy, where the defending nation forms alliances with nations around itself so that they too can attempt to intercept the TBM in its earlier phases as it travels to

the impact location. Therefore, the game is divided into two smaller games: the first being a cooperative multiplayer game devising a set of strategies for the coalition of nations to utilise throughout a TBMs flight path, and the second is the bargaining and cooperation game between the defending nations and potential allies.

The basis of the cooperative game to shoot down the TBM is a strategy called 'shoot-look-shoot.' It relies on a set of N nations attacking the target using a set of strategies - their interceptor missiles - M each of which has its own Probability of Interception P_i. As the TBM flies, each nation n in N will fire its missile(s) m_n to intercept the TBM, then look to see if it has successfully neutralised the threat. If it has failed, the next interceptor m_{n+1} will be fired. The game's problem is then reduced to optimising the probability of interception of the whole set of strategies, such that it has a feasible likelihood of stopping the TBM. Game theory is useful here because the principles of cooperative game theory provide a strong mathematical framework by which an equilibrium solution can be reached for the set of cooperating nations as a whole.

The second game described is based on negotiating with other nations to form an alliance. For these other nations, participation in this game is a risk because it makes them another potential target for the attacking force. To solve this game, the defending nation must accurately evaluate the interceptor cost savings, that is, how much there is to gain by preventing the impact of the TBM. With these savings becoming the surplus that cooperating nations can share in, potential allies then negotiate over how those savings will be shared, in proportion to what interception resources they have to offer.

Shinar and Shima [111] continues the research of both pursuit-evasion games and ballistic missile defence with a zero sum game of a highly maneuverable ballistic missile avoiding an interceptor missile. More specifically, it ties in an imperfect information element to the game, where the ballistic missile knows it is being attacked by anti-missle missiles, but knows little about their trajectory or launching location. In this game, the two players are the ballistic missile and the interceptor. If the ballistic missile uses a pure strategy, it will likely be hit because it either (a) cannot react quickly enough to an opponent it has little information about or (b) will move predictably and allow for a straightforward trajectory towards collision. Therefore, the best solution to the game for the ballistic missile is in mixed strategies.

The mixed strategy will incorporate stochasticity in its flight pattern, assigning a probability distribution over a set of pure strategies. These pure strategies will be based on essential navigational heuristics, which will likely be known or easily discovered by the interceptor. By applying a small number of rapid and random switches in strategy, the ballistic missile can maximise its potential for avoiding interception, and force the complexity of timing calculation back onto the interceptor.

Bogdanovic et al. [112] investigate a target selection problem for multi-target tracking using a game-theoretic perspective. This is an important problem in a multi-function radar network as it needs to perform multiple functions such as volume surveillance and fire control simultaneously while effectively managing the available radar resources to achieve specified objectives. Thus, in effect they tackle a radar resource management problem in [112] and use non-cooperative game-theoretic approaches to find optimal solutions for this problem. They formulate the problem in a framework where each radar is considered to be autonomous; there is no central control engine informing the radars of their optimal strategies nor is there any communication among the radars. First they consider a case where all radars share common interests with respect to the targets and for this problem they propose a distributed algorithm based on the best response dynamics to find the Nash equilibria points. This problem is then extended to a more realistic case of heterogeneous interests among radars and partial target observability. For this case, they employ the solution concept of correlated equilibria and propose an efficient distributed algorithm based on regret matching which is shown to achieve comparable performance to the more computationally intensive centralized approach.

Finally, Parras et al [113] examine a pursuit-evasion game, involving anti-jamming strategies for Unmanned Aerial Vehicles (UAVs). The game operates within a continuous time frame and is therefore dynamic, being solved with the help of differential Game Theory. In somewhat of a culmination of the aforementioned work, it combines elements of communication optimisation, sensor evasion and navigation. Given that UAVs require strong communication for control and the relaying of information, this dependency makes UAVs incredibly susceptible to jamming attacks. There are multiple strategies to both jam and anti-jam these communications, and this can be considered a zero-sum game where the UAV must attempt to optimise its communication capacity. There is usually uncertainty about the positioning and movement of the jamming agent, so the game is an imperfect information differential game. The most important payoff for the UAV is avoiding losing communication to jamming, and it can do this by maneuvering to make approximations for the distance of jamming agents, and thus avoiding them.

3.7. Papers dealing with Homeland Security

The key components of homeland security addressed by game theory are cyber security, modelling terrorism threats and defence contracts. With many applications in computer science [114–116], game theory fits well into cyber security problems. Game theory combines the strict mathematical rigour of computer science, with more psychological and philosophical elements like attacker incentives and mindset, as well as human vulnerabilities in cyber security. Terrorism modelling similarly benefits the human considerations of game theory, since so much of their impact is not easily quantifiable, including social, economic and other spheres affected by terrorist threats, all of which are modled in a game-theoretic setting. These variables are necessary to understand the potential of terrorist threats and attacks, and to formulate strategies to best deal with them. Finally, game theory is suitable for a topic such as contracting and subcontracting because it captures interactions between selfish individuals effectively[1,2].

The paper by Tom Litti [117] provides a brief summary of how traditional network security heuristics can be updated with more precision, and how game theory can help network security engineers design strategies to properly predict, mitigate, and handle threatened networks. He develops a qualitative method for valuing the potential risks and costs of attacks on networks. While a fairly short paper, it does provide some cyber security situational examples of game theory in practice. For example, he models a two player zero sum game to represent an attacker and a security system. The nodes have their own interdependence, vulnerabilities and security assets, but cooperate to minimise the attacker's potential to compromise the system. This review is a great introduction to understanding how cyber security scenarios can be modeled in a game-theoretic framework.

Jhawar et al [118] cover a more specific approach of game theory, namely, Attack-Defence Trees (ADTs) to model scenarios. Here ADTs are used to map potential attack and defence scenarios on a system which is equipped with automatic defence protocols. The system needs to be comprehensive in addressing all possible vulnerabilities, as well as generating responses which adapt to the aggressively evolving situation of cyber security attacks. Currently, ADTs only provide upfront system analyses. Having a reactive strategy to cyber security is important because Attackers constantly change their attack strategies for offence, and so the split second judgements that are made in response can make the difference between a successful and a failed defence of a system.

In [118] they model a simple game of Attacker and Defender—a hacker and a security network administrator. The hacker attempts to breach the integrity of the system, and for each move they make, the administrator will have devised a reactive strategy based on the attacker's attempt. The greatest utility of this method comes from the ability to convert long extensive form games into a graphical layout for easier understanding and communication.

Gonzalez [119] clearly outlines a standard two player competitive game of attacker and defender, and then addresses two aspects—Instance Based Learning Theory and Behavioural game theory. The former compiles cognitive information into a representation known as an instance. Each instance has a three part structure of situation, decision and utility—a standard game. However, it is the interaction between instances which is critical to this approach. Instance Based Learning Theory uses the learning from the outcome of each instance to feedback into the situation of the next instance, hopefully leading to better decisions in later iterations. This is notably similar to the reinforcement learning technique in machine learning. On the other hand, Behavioural game theory involves devising a strategy where we assess a variety of factors, to make more precise long term evaluations of targets and resources so that utility scores more closely reflect real life value. Once again, game theory allows us to draw upon social information in cyber security situations, and assess how that will affect behaviour for both players in the game. Other such factors include motivational factors of the players, completeness of information for each player, technological constraints and inefficiencies between player and technology. Gonzalez stresses that the importance of accommodating for these factors in any cyber security model will help make more realistic and useful policies for cyber defence.

One common use of cyber security is in the prevention of terrorism. Hausken et al [120] cover both terrorism and natural disaster modelling with some guiding game theory principles. Terrorism and Natural Disasters are addressed by defending with anti-terrorism, anti-disaster, and anti-all-hazard investments. Making projections about the likelihoods of each of these occurrences, defenders must make strategic decisions towards the amount of investment dedicated to each defence. Costs to be considered in the utility function of each of these situations include: the intelligence of terrorists or the randomness/ environmental control of natural disasters; the intensity of an attack/ disaster; and differences in evaluation of target value between a terrorist and defender. The game-theoretic approach used in this analysis captures the defender's effort in combating each threat. Depending on the likelihood of each event, combined with the costs of each defensive system, defenders can derive an optimal division of funds.

Kanturska et al. [121] write a rigorous inspection of how transport network reliability can be assessed using game theory, when attack probability on different locations is unknown. The approach favoured using a minimax algorithm to distribute risk across multiple paths as long as the travel costs are small relative to potential losses incurred by an attack. This would be useful in negating the potential risk associated with safely escorting a VIP through a city. Game theory is helpful in this situation because it can analyse network reliability when the attack probabilities are unknown.

Bier [122] presents useful game theory based suggestions for policy insights and investment decisions, insurance policy premiums and more. Her work discusses the weakest link model: a strategy focusing all resources on preventing the worst utility scenario. This is generally weak in practice, and she suggests instead to consider hedging those investments with a variety of defense strategies for different potential targets. The paper looks into standard terrorist/ defender games, and how security investments can change the landscape of attacker-defender interactions for the whole community. This is mainly done through its own scoping study, with one of the key takeaways being that terrorism mitigation systems can benefit from game theory because it adds an extra layer of consideration of the terrorists' response to any defence mechanisms. Hence, game theory, combined with the holistic approach of risk and reliability analysis over all systems, can provide a more comprehensive assessment of all the potential risks and vulnerabilities in counter terrorism strategies.

Cioaca [123] investigates a similar question to Bier et al [84] as mentioned earlier, but specifically towards aviation security. For Cioaca, the problem is summarised by targeting both the cost of security measures for airports and the cost of maintaining a stable and resilient system of defence. Key strategies are: preventing the attack or threat entirely (perhaps by removing all access to the target location or restricting airlines permissions if they fail to adhere to guidelines to helping); managing the temporal dimension of

the attack (the length of the attack and the subsequent time to recover from it); understanding all direct and indirect losses (both casualties and related damages like contamination or infection, compromised secondary security measures, or reputational/ signal ramifications); and the costs of mitigation, response and recovery.

The model is built around several factors and parameters. The first and most critical is human losses and material damages $C(H, D|b_t)$. The most obvious and direct casualties of an attack, these two losses are highly negative payoffs in such an attack, and are often greater than any cost to prevent them. Human losses H are hard to numerically quantify, and therefore when making proper assessments about resource division, understanding how to minimise human losses across different groups of humans and in different dimensions is one of the most difficult aspects of this problem. Material damages D can be monetarily quantified, but often the run-on effects of such damages are where the significant losses are incurred. These losses can lead to total infrastructure shutdowns, ceasing operation of the facility, loss of jobs of workers and even potentially the slow decline and shutdown of the facility altogether. The second major factor in this game is the budget allocated to security systems b_t. Organisational and managing bodies will only have a certain amount of resources allocated to a security system T. The next factor is the number s of security system components as this will be how the budget is composed. Each of these components are partitioned into one of n separate system sub-components. These components are divided amongst a number of targets t_{ik}, and each one of these targets is assigned a probability of being attacked $p_{a_j}^{t_k}(b_{t_k})$ and a value $w_{t_{ik}}$. This can be formally expressed as:

$$\min C(H, D|b_t) = \sum_{j=1}^{l} \sum_{i=1}^{n} \sum_{k=1}^{s} p_{a_j}^{t_{ik}} \cdot p_{v_j}^{t_{ik}} \cdot w_{t_{ik}} \tag{6}$$

For any system of resource division, Ciaoca advocates establishing dimensions of measuring system resilience. This is divided into static resilience, the efficient allocation of resources; and dynamic resilience, recovery speed of the system after the shock, including long term investment in-flows. These two forms of resilience signify the strength of a system both before, during and after an attack. With respect to game theory, Ciaoca's study defines a game clearly, and incorporates a myriad of complex and interconnected parameters to outline an effective and calculable game model.

The final paper we discuss on home land security is by Gardener and Moffat [124]. A distinct and unique approach to game theory in defence, this paper covers the notion of developing a strategy to assess defence contractors and their potential performance/ ability to meet contractual obligations. In game theory terms, this problem is a question of cooperation and defection. It suggests quantitative methods through which defence departments can more rigorously assess contracts and bidding scenarios, and therefore wisely select contractors, and protect their budget. Gardener and Moffat further the understanding of change requirements in project management at different bidding stages of defence acquisition projects. The factor they focus on is the *conspiracy of optimism*, whereby projects spiral out of control—past budget limits and necessary deadlines—due to irrational expectations of project progress. Often this 'conspiracy' is a drive towards making short term gains, and in fact leads to overall losses. The bidding game that is played becomes less about the success of the project, and more about profit capitalisation. The paper presents a guided approach to understanding how 'cooperative' games still lead to selfish motivations. Such is the unreliability of contract outcomes that the game further degenerates to a two player game against the industry of contractors as a whole.

3.8. Papers dealing with Other / Mixed warfare

Zhang and Meherjerdi [125] investigate how groups of multiple unmanned vehicles can be used and controlled using game-theoretic methods in different communication frameworks. Dividing a mission for a single unmanned vehicle amongst multiple unmanned vehicles yields more effective task allocation and performance. The separation of labour away from one powerful single vehicle to several smaller vehicles provides flexibility, adaptability and improved fault tolerance. The uses of such a network are surveillance, exploration, satellite clustering, combining Unmanned Underwater Vehicles (UUVs) and submarines, planes and UAVs, and cooperative robot reconnaissance. As evident from this list, the strategy is incredibly powerful because if its ability to combine resources across multiple domains. To be able to operate a cohesive unit of resources across multiple unmanned systems for combat or exploration brings an unprecedented level of information and control which would accelerate military performance tremendously.

Search is a 'hide and seek' game with a long history in military applications [126–135]. The theory was pioneered by Koopman [126] primarily in the military context (search for an escaping target), followed by developments by Stone *at al* [127]. The applications include submarine hunting, mine detection, rescue operations, risk for the first responders, and backtracking of a hazardous source [126–135]. This framework provides the optimal *a priori* search plan for a given detection model, target motion and the cost of the search. The cost of the search may include time of the search, probability of escape (for a target), exposure risk (for a searcher), information entropy, or situation awareness (map of probability of target location). The searcher can be a moving platforms (UAV, UV, patrolling boats, helicopters, robots, people) and the targets can be static, movable, blind, silent, or emitting. In this context the simultaneous localisation and mapping (SLAM) algorithms are often used [136].

The new direction of this research (inspired by some biological applications) employs the ideas of infotaxis [137], or real-time control of the searchers movement based on information (entropy) gain extracted from the environment (sporadic measurements, forbidden areas, communication between searchers). This method resembles analogy with navigation of some animals [137,138].

4. Classification of papers

In the previous section, it would have been noticeable that many papers have applicability in more than one domain, and use myriad types of games and model a range of players. It is therefore imperative to classify the reviewed papers in a principled manner. To do so, we use the classification scheme already introduced in Table 1 in section 2.

In particular, the reviewed papers could be classified based on (1) the domain or type of warfare (2) the type of game or games used in the paper (3) the nature of players modelled in the paper. The domain can be broadly classified into Traditional (T) or Modern (M), and more specifically into Land warfare, Naval warfare, Aerial warfare, Cyber warfare and Space Warfare. The type of games used can also have a complex classification, based on whether the games were Non-cooperative or cooperative, sequential or simultaneous, discrete or continuous, zero sum or non-zero sum. Finally, the games can be two-player, three-player, or multi-player games. All of this is succinctly captured in Table 1.

In Table 2, we provide a self-explanatory, elaborate classification of all the reviewed papers based on the above-mentioned classification scheme.

Table 2. Defence Game Theory Classifed using Table 1

Title	Authors	Classification Code using Military Defence Game Theory Classification System
Game Theoretic analysis of adaptive radar jamming	Bachmann et al	T-N-IW-NCo-Sim-D-ZS-
Target selection for tracking in multifunction radar networks: Nash and Correlated equilibria	Bogdanovic et al	T-A-IW-NCo-Sim-D-ZS-
Power allocation game between a radar network and multiple jammers	Deligiannis et al	T-A-IW-NCo-Sim-D-ZS-
Strategies for defending a coastline against multiple attackers	Garcia et al	T-A-IW-NCo-Sim-D-ZS-
A game theory approach to target tracking in sensor networks	Gu et al	T-L-IW-NCo-Sim-D-ZS-
Joint Power allocation and beamforming between a multistaticradar and jammer based on game theory	He et al	T-A-IW-NCo-Sim-D-ZS-
Game theoretic situation and transmission in unattended ground sensor networks: a correlated equilibrium approach	Krishnamurthy et al	T-L-(RAW & AMW)-NCo-Sim-D-ZS-
Network enabled missile deflection: games and correlated equilibrium	Maskery et al	T-N-(RAW & IW)-Co-Sim-D-NZS-
Decentralised algorithms for netcentric Force Protection against anti-ship missiles	Maskery et al	T–N-(RAW & IW)-NCo-Sim-D-ZS-
Stochastic game approach to air operation	McKeaney et al	T-A-RAW-NCo-Sim-D-ZS-
Hybrid Game Theory and D-S Evidence Approach to Multiple UCAVs Cooperative Air Combat Decision	Xingxing Wie et al	T-A-RAW-NCo-Sim-D-ZS-
Cooperative Occupancy Decision Making of Multiple-UAV in Beyond-Visual Range Air-Combat: A Game Theory Approach	Yingying Ma et al	T-A-RAW-NCo-Slt-D-ZS-
A Simple Game-Theoretic Approach to Suppression of Enemy Defenses and Other Time Critical Target Analyses	Thomas Hamilton et al	T-A-RAW-NCo-Slt-D-ZS-
The Cooperative Ballistic Missile Defence Game	Lanah Evers et al	T-A-RAW-NCo-Slt-D-ZS-
Cooperative Missile Guidance for Active Defense of Air Vehicles	Eloy Garcia et al	T-A-RAW-NCo-Slt-D-ZS-
Counterfactual regret minimization for integrated cyber and air defense resource allocation	Andrew Keith et al	M-C-RAW-NCo-Slt-D-ZS-
Game Theoretic Model for the Optimal Disposition of Integrated Air Defense System Assets	Chan Y. Han et al	T-A-RAW-NCo-Seq-D-ZS-
Differential game theory and applications to Missile Guidance	Faruqi	T-A-WAW-NCo-Slt-C-ZS-(2P,3P,NP)
Assessment of Aerial Combat Game via Optimisation-Based Horizon Control	Baspınar, Barıs et al	T-A-RAW-NCo-Slt-D-ZS-
Target Differential Game with Two Defenders	David Casbeer	T-A-WAW-NCo-Slt-C-ZS-3P
The optimal search for weak targets	Koopman, Stone et al	T-N-AMW-NCo-Slt-D-ZS-
Optimal Strategy for Target Protection with a defender in the pursuit-evasion scenario	Qilong et al	

Differential game theory with applications to missiles and autonomous systems guidance	Faruqi	T-A-(AMW,WCW)-NCo-Slt-C-ZS-
A game theoretical interceptor guidance law for ballistic missile defence.	Shinar et al	T-A-(AMW,WCW)-NCo-Slt-C-ZS-
Pursuit-Evasion games: a tractable framework for anti-jamming in aerial attacks	Parras et al	T-A-AMW-Nco-Slt-C-ZS-
A simple game theoretic approach to suppression of enemy defences and other time-critical target analyses	Hamilton et al	T-A-(RAW,WCW)-NCo-Slt-D-ZS-
Choosing What to Protect: Strategic Defence allocation against an unknown attacker	Bier et al	T-L-RAW-NCo-Slt-D-ZS-
Considerations on Optimal Resource allocation in avation security	Cioaca	T-L-RAW-NCo-Slt-D-ZS-
Horsemen and the empty city: A game theoretic examination of deception in Chinese military legend	Cotten et al	T-L-IW-NCo-Slt-D-ZS-
An Economic Theory of Destabilisation War	Gries et al	T-L-IW-NCo-Seq-D-ZS-
Game theoretic approach towards network security: A review	Tom Litti	M-C-IW-NCo-Slt-D-ZS-
Automating cyber defence responses using attack-defence trees and game theory	Jhawar et al	M-C-WCW-NCo-Slt-D-ZS-
From individual decisions from experience to behavioural game theory	Gonzalez	M-C-IW-NCo-Slt-D-ZS-
Game Theoretic Approaches to Attack Surface Shifting. Moving Target Defense II	Manadhata	
Improving reliability through Multi-Path routing and Link Defence: An Application of Game Theory to Transport	Kanturska et al	M-C-WCW-NCo-Slt-D-ZS-
Game Theoretic and Reliability models in counter-terrorism and security	Bier et al	M-C-(IW,AMW,WCW)-NCo-Slt-D-ZS-
Changing behaviours in defence acqusition: a game theory approach	Gardener et al	
Joint Transmit Power and Bandwidth Allocation for Cognitive Satellite Network based on Bargaining Game Theory	Zhong et al	M-S-RAW-Co-Slt-D-NZS
The Research and Simulation of a Satellite Routing Algorithm based on Game Theory	Qiao et al	M-S-WCW-NCo-Slt-D-ZS
A survey of multiple unmanned vehicles formation control and coordiation. Normal and fault situations	Zhang et al	T-A-WCW-Co-Slt-D-NZS

5. Opportunities for Further Research

The reviewed papers haver shown that game theory can provide a unifying framework to analyse decision making behaviours of agents in defence contexts. In this section, we briefly discuss a range of potential defence scenarios where game theory has hitherto not been applied, but would make a useful contribution if applied in future.

A recent investigation by Defence Advanced Research Projects Agency (DARPA) into 'Mosaic warfare' [139] is an example of such a potential future application of game theory. The idea is mentioned in Zhang [125] in the context of operating multiple unmanned vehicles, and proposes having a lot of smaller

cost-efficient resources interconnected in a 'mosaic' network, such that if several units are destroyed, the overall nature of the network stands, much like how a mosaic retains its image even if a few tiles are removed. The goal is that such a vast array of resources with different capabilities will be able to overwhelm the enemy with its completeness and complexity. It utilises principles of concurrency to address the intricacy of the connections in systems of millions of sensors and actuators. These systems in turn must handle inter-system communication. If successfully implemented, such a system of systems can provide military strategists with an overwhelmingly powerful network of weaponry and resources, which can defeat opponents with the sheer scale and complexity of their dynamics. This method of combining different parts of the arsenal maximises the benefits of each, and reintroduces a focus on expendability of resources, rather than focussing on a few pieces of high value weaponry. This in turn builds resilience and adaptability into the strategy, a shift away from heavyweight, single focus attack methods. Since there are a high number of lower cost resources which need to cooperate for the best outcome, this scenario could be modelled as a multi-player co-operative game at one level, while the strife with the opponent(s) can be modelled as a muti-player non-cooperative game. it cane be noted that the concept of 'Mosaic warfare' is essentially similar to the more general concept of agent-based modelling, which has already been used in several diverse contexts, ranging from Ageless Aerospace Vehicle design [140,141] to modelling of infectious disease dynamics [142], and game theory has already been used successfully in some of these contexts [143,144].

Another area, within the context of naval warfare, where game theory could be fruitfully applied in naval susceptibility. in analysing naval susceptibility, naval vessels factor in their environment, movement patterns and potential adversary sensors to calculate their risk of detection when moving covertly [145]. Such an application has overlaps between commonly studied tracking problems in defence science, as explained in Gu's [107] which describes tracking using sensor networks. Such a scenario, as elaborated before, could be modelled as a two-player non-cooperative differential game, with detection being the main payoff parameter for each player.

Indeed, ground based tracking problems could benefit from the application of game theory as well, and papers in this area so far has been few. Ground based tracking problems may appear in both ground based military applications (classified here as ground warfare), as well as domestic security and anti-terrorism applications (classified here as home-front warfare), where the the ability of security agencies to track individuals' movements throughout a society - including their locations, their social network and their motivations - is a crucial capability[146]. The later scenario could be modelled as a two player game of pursuit and evasion, or perhaps just pursuit and reconnaissance with the aim of not revealing the pursuit to the target, while the target would try to identify pursuit. The amount of predictive information gained from covert tracking would be the payoff in this situation.

Modelling cyber warfare is another area where game theory could be applicable, and again there have been few papers addressing this niche, other than papers coming primarily from the computer science domain and focus primarily on cyber security. Kim et al [147] describe cyber warfare scenarios which are integral to all military operations, and highlight the critical role played by new technological paradigms such as Internet of Things (IoT) and Brain Computer Interfaces. Defence experts increasingly have a need to predict and preempt cyber warfare strategies of hostile players. Modelling decision making scenarios involving Cyber warfare scenarios with novel technological interfaces is an area where game theory can play a vital role.

6. Conclusions

Game theory has proven itself as a versatile and powerful tool for obtaining vital insights about the decision making processes of agents and players in a number of fields. In this review paper, we elaborated

several scenarios in which game theory could be applied in defence science and technology, and presented a succinct review of existing research in this direction. We introduced an extensive classification for the reviewed papers, based on the kind of warfare studied, type of games used, and the nature of players. Based on the observations made, we identified gaps in the literature where game theory has not been applied extensively so far but has great potential to be applied fruitfully in future, and we discussed the potential directions in which defence applications of game theory could expand in the future

The domain based classification was the primary mode of classification that employed, and in this context we grouped the reviewed papers into papers about tracking systems, aerial warfare, ground warfare, home front warfare and space/ other warfare. For each paper considered, the number and roles of players and game types were defined, and where possible, strategies and payoff functions were discussed. The goal of this exercise was to identify the most commonly analysed domains as well as frequently used game types, and use this knowledge to identify gaps in literature and cross-pollinate ideas across various domains and types of warfare within the defence context.

As the world deals with emerging challenges to peace and stability, the future of humanity depends on our ability to solve problems peacefully. While this is a lofty goal to achieve, the projection of power is decidedly better than an actual armed conflict which will be very costly at many levels, and game theory could indeed play a part in deciding some of the 'soft conflicts' which could play out in the coming years and decades. As the focus on defence strategies and capabilities are likely to increase in the coming years, game theory can serve as an additional tool that defence scientists can use at many levels of abstraction to solve deployment, sensing, tracking, and resource allocation problems.

Author Contributions: MP, SA, and AS conceptualised the research. All authors undertook the literature survey and wrote the manuscript.

Acknowledgments: The authors thank Dr. Farhan Faruqui from the Defence Science and Technology Group, Australia for many fruitful discussions

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Piraveenan, M. Applications of Game Theory in Project Management: A Structured Review and Analysis. *Mathematics* **2019**, *7*, 858.
2. Peldschus, F. Experience of the game theory application in construction management. *Technological and Economic Development of Economy* **2008**, *14*, 531–545.
3. Schotter, A. The economic theory of social institutions **2008**.
4. Morrow, J.D. *Game theory for political scientists*; Princeton University Press, 1994.
5. Hammerstein, P.; Selten, R. Game theory and evolutionary biology. *Handbook of game theory with economic applications* **1994**, *2*, 929–993.
6. Von Neumann, J.; Morgenstern, O. Game theory and economic behavior. *Joh Wiley and Sons, New York* **1944**.
7. Nash, J.F.; others. Equilibrium points in n-person games. *Proceedings of the national academy of sciences* **1950**, *36*, 48–49.
8. Nash, J. Non-cooperative games. *Annals of mathematics* **1951**, pp. 286–295.
9. Martin, B. The selective usefulness of game theory. *Social Studies of Science* **1978**, *8*, 85–110.
10. Machiavelli, N. *The prince (1513)*; 1984.
11. Lattimore, R. *The Iliad of Homer*; 1962.
12. Tzu, S. The art of war. In *Strategic Studies*; Routledge, 2014; pp. 86–110.
13. Myerson, R.B. *Game theory*; Harvard university press, 2013.
14. Osborne, M.J.; others. *An introduction to game theory*; Vol. 3, Oxford university press New York, 2004.
15. Haywood Jr, O. Military decision and game theory. *Journal of the Operations Research Society of America* **1954**, *2*, 365–385.

16. Thunholm, P. Planning under time pressure: An attempt toward a prescriptive model of military tactical decision making. *How professionals make decisions* **2005**, pp. 43–56.
17. Andrzejewski, S. *Military organization and society*; Routledge, 2013.
18. Janowitz, M. The military in the political development of new nations. *Bulletin of the Atomic Scientists* **1964**, *20*, 6–10.
19. DeGrasse, R.W. *Military expansion, economic decline: Impact of military spending on United States economic performance*; Routledge, 2016.
20. Roland, A. Technology, Ground Warfare, and Strategy: The Paradox of American Experience. *The Journal of Military History* **1991**, *55*, 447–468.
21. Paxson, E.W.; Weiner, M.G.; Wise, R.A. Interactions Between Tactics and Technology in Ground Warfare. Technical report, RAND CORP SANTA MONICA CA, 1979.
22. Edmonds, M. Land Warfare. In *International Security in the Modern World*; Springer, 1992; pp. 179–206.
23. Harding, R. *Seapower and naval warfare, 1650-1830*; Routledge, 2002.
24. Speller, I. *Understanding naval warfare*; Routledge, 2018.
25. Paris, M. *Winged warfare: the literature and theory of aerial warfare in Britain, 1859-1917*; Manchester University Press, 1992.
26. Colby, E. Laws of Aerial Warfare. *Minn. L. Rev.* **1925**, *10*, 207.
27. Garner, J.W. Proposed rules for the regulation of aerial warfare. *American Journal of International Law* **1924**, *18*, 56–81.
28. Gingras, J.L.; Ruby, T.Z. Morality in Modern Aerial Warfare. Technical report, AIR UNIV MAXWELL AFB AL, 2000.
29. Dunn, D.H. Drones: disembodied aerial warfare and the unarticulated threat. *International Affairs* **2013**, *89*, 1237–1246.
30. Stoyen, A.D. Method and system for intelligent agent decision making for tactical aerial warfare, 2008. US Patent 7,330,844.
31. Klein, J.J. *Space warfare: Strategy, principles and policy*; Routledge, 2012.
32. Wortzel, L.M. The Chinese People's Liberation Army and Space Warfare. *Astropolitics* **2008**, *6*, 112–137.
33. Johnson-Freese, J. *Space warfare in the 21st century: Arming the heavens*; Routledge, 2016.
34. Blackman, S.S. Multiple-target tracking with radar applications. *Dedham, MA, Artech House, Inc., 1986, 463 p.* **1986**.
35. Lipton, A.J.; Fujiyoshi, H.; Patil, R.S. Moving target classification and tracking from real-time video. Proceedings Fourth IEEE Workshop on Applications of Computer Vision. WACV'98 (Cat. No. 98EX201). IEEE, 1998, pp. 8–14.
36. Li, X.R.; Jilkov, V.P. Survey of maneuvering target tracking. Part I. Dynamic models. *IEEE Transactions on aerospace and electronic systems* **2003**, *39*, 1333–1364.
37. Kasthurirathna, D.; Piraveenan, M.; Harré, M. Influence of topology in the evolution of coordination in complex networks under information diffusion constraints. *The European Physical Journal B* **2014**, *87*, 3.
38. Bar-Shalom, Y.; Li, X.R. Estimation and tracking: Principles, techniques, and software [reviews and abstracts]. *IEEE Antennas and Propagation Magazine* **1996**, *38*, 62.
39. Gandhi, R.; Sharma, A.; Mahoney, W.; Sousan, W.; Zhu, Q.; Laplante, P. Dimensions of cyber-attacks: Cultural, social, economic, and political. *IEEE Technology and Society Magazine* **2011**, *30*, 28–38.
40. Harris, J. War and social history: Britain and the home front during the Second World War. *Contemporary European History* **1992**, *1*, 17–35.
41. Maltby, S.; Keeble, R. *Communicating war: Memory, military and media*; Arima Publishing, 2007.
42. Barough, A.S.; Shoubi, M.V.; Skardi, M.J.E. Application of Game Theory Approach in Solving the Construction Project Conflicts. *Procedia - Social and Behavioral Sciences* **2012**, *58*, 1586 – 1593. 8th International Strategic Management Conference, doi:http://dx.doi.org/10.1016/j.sbspro.2012.09.1145.
43. Kasthurirathna, D.; Piraveenan, M. Topological stability of evolutionarily unstable strategies. 2014 IEEE Symposium on Evolving and Autonomous Learning Systems (EALS). IEEE, 2014, pp. 35–42.

44. Myerson, R. Game Theory: Analysis of Conflict Harvard Univ. *Press, Cambridge* **1991**.
45. Von Neumann, J. *On the theory of games of strategy*; Princeton University Press Princeton, NJ, 1928.
46. Von Neumann, J.; Morgenstern, O. Theory of Games and Economic Behavior.(1967 edition) John Wiley and Sons. *Inc., New York, New York* **1944**, *641*.
47. Rasmusen, E.; Blackwell, B. Games and information. *Cambridge, MA* **1994**, *15*.
48. Kasthurirathna, D.; Piraveenan, M. Emergence of scale-free characteristics in socio-ecological systems with bounded rationality. *Scientific reports* **2015**, *5*.
49. Kasthurirathna, D.; Piraveenan, M.; Harre, M. Evolution of coordination in scale-free and small world networks under information diffusion constraints. Advances in Social Networks Analysis and Mining (ASONAM), 2013 IEEE/ACM International Conference on. IEEE, 2013, pp. 183–189.
50. Thedchanamoorthy, G.; Piraveenan, M.; Uddin, S.; Senanayake, U. Influence of vaccination strategies and topology on the herd immunity of complex networks. *Social Network Analysis and Mining* **2014**, *4*, 1–16.
51. Thakor, A.V. Game theory in finance. *Financial Management* **1991**, pp. 71–94.
52. Bshary, R.; Oliveira, R.F. Cooperation in animals: toward a game theory within the framework of social competence. *Current Opinion in Behavioral Sciences* **2015**, *3*, 31–37.
53. Hadjichrysanthou, C.; Broom, M. When should animals share food? Game theory applied to kleptoparasitic populations with food sharing. *Behavioral ecology* **2012**, *23*, 977–991.
54. Lundberg, S.; Pollak, R.A. Noncooperative bargaining models of marriage. *The American Economic Review* **1994**, *84*, 132–137.
55. Manser, M.; Brown, M. Marriage and household decision-making: A bargaining analysis. *International economic review* **1980**, pp. 31–44.
56. Grossbard-Shechtman, S. Marriage market models. *The new economics of human behavior* **1995**, pp. 92–112.
57. Murray, A.G. A game theory based framework for assessing incentives for local area collaboration with an application to Scottish salmon farming. *Preventive veterinary medicine* **2014**, *115*, 255–262.
58. Gibbons, R.S. *Game theory for applied economists*; Princeton University Press, 1992.
59. Shubik, M. Game theory, behavior, and the paradox of the prisoner's dilemma: Three solutions. *Journal of Conflict Resolution* **1970**, *14*, 181–193.
60. Polak, B. Econ 159: Game theory. *Open Yale courses, Department of Economics, http://oyc. yale. edu/economics* **2007**.
61. Kasthurirathna, D.; Nguyen, H.; Piraveenan, M.; Uddin, S.; Senanayake, U. Optimisation of strategy placements for public good in complex networks. Proceedings of the 2014 International Conference on Social Computing. ACM, 2014, p. 1.
62. Kasthurirathna, D.; Piraveenan, M.; Uddin, S. Evolutionary stable strategies in networked games: the influence of topology. *Journal of Artificial Intelligence and Soft Computing Research* **2015**, *5*, 83–95.
63. Binmore, K. *Playing for real: a text on game theory*; Oxford university press, 2007.
64. Hart, S. Games in extensive and strategic forms. *Handbook of game theory with economic applications* **1992**, *1*, 19–40.
65. Ritzberger, K.; others. Foundations of non-cooperative game theory. *OUP Catalogue* **2002**.
66. Peleg, B.; Sudhölter, P. *Introduction to the theory of cooperative games*; Vol. 34, Springer Science & Business Media, 2007.
67. Bell, M.; Perera, S.; Piraveenan, M.; Bliemer, M.; Latty, T.; Reid, C. Network growth models: A behavioural basis for attachment proportional to fitness. *Scientific reports* **2017**, *7*, 42431.
68. Kuhn, H.; Arrow, K.; Tucker, A. *Contributions to the Theory of Games*; Number v. 2 in Annals of Mathematics Studies, Princeton University Press, 1953.
69. Kamien, M.I.; Schwartz, N.L. *Dynamic optimization: the calculus of variations and optimal control in economics and management*; Courier Corporation, 2012.
70. Von Stackelberg, H. *Market structure and equilibrium*; Springer Science & Business Media, 2010.
71. Calcagno, R.; Kamada, Y.; Lovo, S.; Sugaya, T. Asynchronicity and coordination in common and opposing interest games. *Theoretical Economics* **2014**, *9*, 409–434.
72. Schelling, T.C. The strategy of conflict. Prospectus for a reorientation of game theory. *Journal of Conflict Resolution* **1958**, *2*, 203–264.

73. Schelling, T.C. *The strategy of conflict*; Harvard university press, 1980.
74. Lewis, D. *Convention: A philosophical study*; John Wiley & Sons, 2008.
75. Nash Jr, J.F. The bargaining problem. *Econometrica: Journal of the Econometric Society* **1950**, pp. 155–162.
76. Myerson, R.B. Nash equilibrium and the history of economic theory. *Journal of Economic Literature* **1999**, *37*, 1067–1082.
77. Rubinstein, A. Perfect equilibrium in a bargaining model. *Econometrica: Journal of the Econometric Society* **1982**, pp. 97–109.
78. Camerer, C.F. *Behavioral game theory: Experiments in strategic interaction*; Princeton University Press, 2011.
79. Aumann, R.J.; Hart, S. *Handbook of game theory with economic applications*; Vol. 1, North-Holland Amsterdam, 1992.
80. Aumann, R. Game theory, The New Palgrave Dictionary of Economics (New York)(SN Durlauf and LE Blume, eds.), 2008.
81. Smith, J.M. *Evolution and the Theory of Games*; Springer, 1993.
82. Le, S.; Boyd, R. Evolutionary dynamics of the continuous iterated Prisoner's dilemma. *Journal of theoretical biology* **2007**, *245*, 258–267.
83. Cho, I.K.; Kreps, D.M. Signaling games and stable equilibria. *The Quarterly Journal of Economics* **1987**, *102*, 179–221.
84. Bier, V.; Oliveros, S.; Samuelson, L. Choosing What to Protect: Strategic Defensive Allocation against an Unknown Attacker. *Journal of Public Economic Theory* **2007-08**, *9*, 563,587.
85. Cotton, C.; Liu, C. 100 Horsemen and the empty city: A game theoretic examination of deception in Chinese military legend. *Journal of peace research* **2011**, *48*, 217–223.
86. Gries, T.; Haake, C.J.; others. An Economic Theory of'Destabilization War'. Technical report, Paderborn University, CIE Center for International Economics, 2016.
87. Krishnamurthy, V.; Maskery, M.; Ngo, M.H. Game theoretic activation and transmission scheduling in unattended ground sensor networks: A correlated equilibrium approach. In *Wireless Sensor Networks: Signal Processing and Communications Perspectives*; Wiley, 2007; pp. 349–388.
88. Levine, A.J. *The pacific war: Japan versus the allies*; Vol. 130, Greenwood Publishing Group, 1995.
89. Maskery, M.; Krishnamurthy, V. Network-enabled missile deflection: Games and correlation equilibrium. *IEEE Transactions on Aerospace and Electronic Systems* **2007**, *43*, 843–863.
90. Maskery, M.; Krishnamurthy, V.; O'Regan, C. Decentralized algorithms for netcentric force protection against antiship missiles. *IEEE Transactions on Aerospace and Electronic Systems* **2007**, *43*, 1351–1372.
91. Bachmann, D.J.; Evans, R.J.; Moran, B. Game theoretic analysis of adaptive radar jamming. *IEEE Transactions on Aerospace and Electronic Systems* **2011**, *47*, 1081–1100.
92. Hamilton, T.; Mesic, R. A simple game-theoretic approach to suppression of enemy defenses and other time critical target analyses. Technical report, RAND CORP SANTA MONICA CA, 2004.
93. Deligiannis, A.; Rossetti, G.; Panoui, A.; Lambotharan, S.; Chambers, J. Power allocation game between a radar network and multiple jammers. 2016 IEEE Radar Conference (RadarConf). IEEE, 2016, pp. 1–5.
94. Garcia, E.; Von Moll, A.; Casbeer, D.W.; Pachter, M. Strategies for defending a coastline against multiple attackers. 2019 IEEE 58th Conference on Decision and Control (CDC). IEEE, 2019, pp. 7319–7324.
95. He, B.; Su, H. Joint Power Allocation and Beamforming between a Multistatic Radar and Jammer Based on Game Theory. 2019 IEEE 11th International Conference on Communication Software and Networks (ICCSN). IEEE, 2019, pp. 337–341.
96. McEneaney, W.M.; Fitzpatrick, B.G.; Lauko, I.G. Stochastic game approach to air operations. *IEEE Transactions on Aerospace and Electronic Systems* **2004**, *40*, 1191–1216.
97. Wei, X.; Duan, H.; Wang, Y. Hybrid game theory and DS evidence approach to multiple UCAVs cooperative air combat decision. International Symposium on Neural Networks. Springer, 2009, pp. 236–246.
98. Ma, Y.; Wang, G.; Hu, X.; Luo, H.; Lei, X. Cooperative occupancy decision making of Multi-UAV in Beyond-Visual-Range air combat: A game theory approach. *IEEE Access* **2019**, *8*, 11624–11634.

99. Evers, L.; Barros, A.I.; Monsuur, H. The cooperative ballistic missile defence game. International Conference on Decision and Game Theory for Security. Springer, 2013, pp. 85–98.
100. Garcia, E.; Casbeer, D.W.; Fuchs, Z.E.; Pachter, M. Cooperative missile guidance for active defense of air vehicles. *IEEE Transactions on Aerospace and Electronic Systems* **2017**, *54*, 706–721.
101. Han, C.Y.; Lunday, B.J.; Robbins, M.J. A game theoretic model for the optimal location of integrated air defense system missile batteries. *INFORMS Journal on Computing* **2016**, *28*, 405–416.
102. Başpınar, B.; Koyuncu, E. Assessment of aerial combat game via optimization-based receding horizon control. *IEEE Access* **2020**, *8*, 35853–35863.
103. Casbeer, D.W.; Garcia, E.; Pachter, M. The target differential game with two defenders. *Journal of Intelligent & Robotic Systems* **2018**, *89*, 87–106.
104. Keith, A.; Ahner, D. Counterfactual regret minimization for integrated cyber and air defense resource allocation. *European Journal of Operational Research* **2021**, *292*, 95–107.
105. Zhong, X.; Yin, H.; He, Y.; Zhu, H. Joint Transmit Power and Bandwidth Allocation for Cognitive Satellite Network Based on Bargaining Game Theory. *IEEE Access* **2018**, *7*, 6435–6449.
106. Qiao, R.; Zhao, X. The Research and Simulation of the Satellite Network Routing Algorithm Based on Game Theory. 2013 International Conference on Information Science and Cloud Computing. IEEE, 2013-12, pp. 124,130.
107. Gu, D. A game theory approach to target tracking in sensor networks. *IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics)* **2010**, *41*, 2–13.
108. Sun, Q.; Qi, N.; Huo, M. Optimal strategy for target protection with a defender in the pursuit-evasion scenario. *The Journal of Defense Modeling and Simulation: Applications, Methodology, Technology* **2018**, *15*, 154851291875622.
109. Faruqi, F.A. *Differential game theory with applications to missiles and autonomous systems guidance*; John Wiley & Sons, 2017.
110. Evers, L.; Barros, A.I.; Monsuur, H. The Cooperative Ballistic Missile Defence Game. Decision and Game Theory for Security; Das, S.K.; Nita-Rotaru, C.; Kantarcioglu, M., Eds.; Springer International Publishing: Cham, 2013; pp. 85–98.
111. Shinar, J.; Shima, T. A game theoretical interceptor guidance law for ballistic missile defence. Proceedings of 35th IEEE Conference on Decision and Control. IEEE, 1996, Vol. 3, pp. 2780–2785.
112. Bogdanović, N.; Driessen, H.; Yarovoy, A.G. Target selection for tracking in multifunction radar networks: Nash and correlated equilibria. *IEEE Transactions on Aerospace and Electronic Systems* **2018**, *54*, 2448–2462.
113. Parras, J.; Zazo, S.; Del Val, J.; Zazo, J.; Macua, S.V. Pursuit-evasion games: a tractable framework for antijamming games in aerial attacks. *EURASIP Journal on Wireless Communications and Networking* **2017**, *2017*, 69.
114. Halpern, J.Y. Computer science and game theory. *The New Palgrave Dictionary of Economics: Volume 1–8* **2008**, pp. 984–994.
115. Apt, K.R.; Grädel, E. *Lectures in game theory for computer scientists*; Cambridge University Press, 2011.
116. Shoham, Y. Computer science and game theory. *Commun. ACM* **2008**, *51*, 74–79.
117. Tom, L. Game-theoretic approach towards network security: A review. 2015 International Conference on Circuits, Power and Computing Technologies [ICCPCT-2015]. IEEE, 2015, pp. 1–4.
118. Jhawar, R.; Mauw, S.; Zakiuddin, I. Automating cyber defence responses using attack-defence trees and game theory. European Conference on Cyber Warfare and Security. Academic Conferences International Limited, 2016, p. 163.
119. Gonzalez, C. From individual decisions from experience to behavioral game theory: Lessons for cybersecurity. In *Moving target defense II*; Springer, 2013; pp. 73–86.
120. Hausken, K.; Bier, V.M.; Zhuang, J., Defending Against Terrorism, Natural Disaster, and All Hazards. In *Game Theoretic Risk Analysis of Security Threats*; Bier, V.M.M.; Azaiez, M.N., Eds.; Springer US: Boston, MA, 2009; pp. 65–97.

121. Kanturska, U.; Schmöcker, J.D.; Fonzone, A.; Bell, M.G.H., Improving Reliability through Multi-Path Routing and Link Defence: An Application of Game Theory to Transport. In *Game Theoretic Risk Analysis of Security Threats*; Bier, V.M.M.; Azaiez, M.N., Eds.; Springer US: Boston, MA, 2009; pp. 199–227.
122. Bier, V., Game-Theoretic and Reliability Methods in Counterterrorism and Security. In *Statistical Methods in Counterterrorism: Game Theory, Modeling, Syndromic Surveillance, and Biometric Authentication*; Wilson, A.G.; Wilson, G.D.; Olwell, D.H., Eds.; Springer New York: New York, NY, 2006; pp. 23–40.
123. Cioacă, C.; others. Considerations on optimum resource allocation in aviation security. *Journal of Defense Resources Management (JoDRM)* **2013**, *4*, 187–192.
124. Gardener, T.; Moffat, J. Changing behaviours in defence acquisition: a game theory approach. *Journal of the Operational Research Society* **2008**, *59*, 225–230.
125. Zhang, Y.; Mehrjerdi, H. A survey on multiple unmanned vehicles formation control and coordination: Normal and fault situations. 2013 International conference on unmanned aircraft systems (ICUAS). IEEE, 2013, pp. 1087–1096.
126. Stone, L.D. Search and screening: General principles with historical applications (BO Koopman). *SIAM Review* **1981**, *23*, 533–539.
127. Stone, L.D. Optimal search for moving targets. *Advances in Applied Probability* **1980**, *12*, 279–280.
128. Washburn, A.R.; others. Two-person zero-sum games **2014**.
129. Washburn, A.; Karataş, M. Multistatic search theory. *Military Operations Research* **2015**, *20*, 21–38.
130. Jing, T.; Meng, Q.H.; Ishida, H. Recent Progress and Trend of Robot Odor Source Localization. *IEEJ Transactions on Electrical and Electronic Engineering* **2021**.
131. Ristic, B.; Skvortsov, A. Intermittent Information-Driven Multi-Agent Area-Restricted Search. *Entropy* **2020**, *22*, 635.
132. Ristic, B.; Skvortsov, A.; Gunatilaka, A. A study of cognitive strategies for an autonomous search. *Information Fusion* **2016**, *28*, 1–9.
133. Ristic, B.; Gilliam, C.; Moran, W.; Palmer, J.L. Decentralised multi-platform search for a hazardous source in a turbulent flow. *Information Fusion* **2020**, *58*, 13–23.
134. Hutchinson, M.; Liu, C.; Chen, W.H. Information-based search for an atmospheric release using a mobile robot: Algorithm and experiments. *IEEE Transactions on Control Systems Technology* **2018**, *27*, 2388–2402.
135. Park, M.; Oh, H. Cooperative information-driven source search and estimation for multiple agents. *Information Fusion* **2020**, *54*, 72–84.
136. Durrant-Whyte, H.; Bailey, T. Simultaneous localization and mapping: part I. *IEEE robotics & automation magazine* **2006**, *13*, 99–110.
137. Vergassola, M.; Villermaux, E.; Shraiman, B.I. ?Infotaxis? as a strategy for searching without gradients. *Nature* **2007**, *445*, 406–409.
138. Bénichou, O.; Loverdo, C.; Moreau, M.; Voituriez, R. Intermittent search strategies. *Reviews of Modern Physics* **2011**, *83*, 81.
139. DARPA. DARPA Tiles Together a Vision of Mosaic Warfare. https://www.darpa.mil/work-with-us/darpa-tiles-together-a-vision-o Accessed: 2019-09-15.
140. Prokopenko, M.; Piraveenan, M.; Wang, P. On convergence of dynamic cluster formation in multi-agent networks. European Conference on Artificial Life. Springer, 2005, pp. 884–894.
141. Piraveenan, M.; Prokopenko, M.; Wang, P.; Price, D. Towards adaptive clustering in self-monitoring multi-agent networks. International Conference on Knowledge-Based and Intelligent Information and Engineering Systems. Springer, 2005, pp. 796–805.
142. Cliff, O.M.; Harding, N.; Piraveenan, M.; Erten, E.Y.; Gambhir, M.; Prokopenko, M. Investigating spatiotemporal dynamics and synchrony of influenza epidemics in Australia: an agent-based modelling approach. *Simulation Modelling Practice and Theory* **2018**, *87*, 412–431.
143. Chang, S.L.; Piraveenan, M.; Pattison, P.; Prokopenko, M. Game theoretic modelling of infectious disease dynamics and intervention methods: a review. *Journal of biological dynamics* **2020**, *14*, 57–89.

144. Piraveenan, M.; Sawleshwarkar, S.; Walsh, M.; Zablotska, I.; Bhattacharyya, S.; Farooqui, H.H.; Bhatnagar, T.; Karan, A.; Murhekar, M.; Zodpey, S.; others. Optimal governance and implementation of vaccination programmes to contain the COVID-19 pandemic. *Royal Society open science* **2021**, *8*, 210429.
145. Vaitekunas, D.A. IR susceptibility of naval ships using ShipIR/NTCS. Infrared Imaging Systems: Design, Analysis, Modeling, and Testing XXI; Holst, G.C.; Krapels, K.A., Eds. International Society for Optics and Photonics, SPIE, 2010, Vol. 7662, pp. 277 – 284.
146. Zalud, B. Unseen But There: A Place for Covert Security **2010**.
147. Kim, S.K.; Cheon, S.P.; Eom, J.H. A leading cyber warfare strategy according to the evolution of cyber technology after the fourth industrial revolution. *International Journal of Advanced Computer Research* **2019**, *9*, 72–80.

eLIFE
elifesciences.org

RESEARCH ARTICLE

A highly tunable dopaminergic oscillator generates ultradian rhythms of behavioral arousal

Ian D Blum[1,2,4], Lei Zhu[1,2], Luc Moquin[2], Maia V Kokoeva[3], Alain Gratton[1,2], Bruno Giros[1,2,5], Kai-Florian Storch[1,2]*

[1]Department of Psychiatry, McGill University, Montreal, Canada; [2]Douglas Mental Health University Institute, Montreal, Canada; [3]Department of Medicine, McGill University, Montreal, Canada; [4]Integrated Program in Neuroscience, McGill University, Montreal, Canada; [5]INSERM UMR S1130 CNRS, UMR8246, Sorbonne University, Paris, France

Abstract Ultradian (~4 hr) rhythms in locomotor activity that do not depend on the master circadian pacemaker in the suprachiasmatic nucleus have been observed across mammalian species, however, the underlying mechanisms driving these rhythms are unknown. We show that disruption of the dopamine transporter gene lengthens the period of ultradian locomotor rhythms in mice. Period lengthening also results from chemogenetic activation of midbrain dopamine neurons and psychostimulant treatment, while the antipsychotic haloperidol has the opposite effect. We further reveal that striatal dopamine levels fluctuate in synchrony with ultradian activity cycles and that dopaminergic tone strongly predicts ultradian period. Our data indicate that an arousal regulating, dopaminergic ultradian oscillator (DUO) operates in the mammalian brain, which normally cycles in harmony with the circadian clock, but can desynchronize when dopamine tone is elevated, thereby producing aberrant patterns of arousal which are strikingly similar to perturbed sleep-wake cycles comorbid with psychopathology.
DOI: 10.7554/eLife.05105.001

***For correspondence:** florian.storch@mcgill.ca

Competing interests: The authors declare that no competing interests exist.

Funding: See page 19

Received: 10 October 2014
Accepted: 28 December 2014
Published: 29 December 2014

Reviewing editor: Richard D Palmiter, Howard Hughes Medical Institute, University of Washington, United States

Copyright Blum et al. This article is distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use and redistribution provided that the original author and source are credited.

Introduction

Ultradian rhythms with periods ranging from one to several hours have been linked to various aspects of mammalian physiology. Usually superimposed on the 24-hr diurnal or circadian rhythm, ultradian oscillations have been observed in the context of locomotion, sleep, feeding, body temperature, serum hormones, and brain monoamines in species ranging from fruit flies to humans (***Tannenbaum and Martin, 1976***; ***Ibuka et al., 1977***; ***Rusak, 1977***; ***Daan and Slopsema, 1978***; ***Honma and Hiroshige, 1978***; ***Dowse et al., 1987***; ***Rivkees, 2003***; ***van Oort et al., 2007***; ***Dowse et al., 2010***; ***Seki and Tanimura, 2014***). These physiological cycles most frequently exhibit periods of 2–6 hr, adopting harmonics of the 24-hr daily light–dark cycle or the endogenous circadian rhythm, when external timing cues are absent. However, the generation of such ultradian rhythms does not depend on a functional circadian system nor a light:dark cycle. Ultradian locomotor oscillations persist in rodents housed in constant darkness even upon ablation of the suprachiasmatic nucleus (SCN), the site of the master circadian pacemaker (***Ibuka et al., 1977***; ***Rusak, 1977***), or genetic disruption of the circadian clock (***Vitaterna et al., 1994***; ***Bunger et al., 2000***) (***Figure 1A,B***). These ultradian activity cycles may not simply be driven by metabolic demand since the 2- to 3-hr rhythm in foraging activity observed in the common vole persists even in the absence of food (***Gerkema and van der Leest, 1991***). Studies in this species further indicate that one adaptive value of ultradian activity rhythms may lie in the facilitation of social synchrony, which is suggested to reduce predator risk in this species

eLife digest The sleep-wake cycle of mammals is controlled by a 'circadian clock' within the brain, which is synchronized to the day–night cycle. However, other aspects of mammalian physiology including alertness and activity levels, as well as appetite and body temperature—fluctuate in cycles that repeat every few hours. These cycles are known as ultradian rhythms, and they may offer survival benefits by enabling potentially risky behaviors, such as foraging, to be coordinated between members of a group.

Despite their widespread nature and the fact that they appear to be conserved in evolution, virtually nothing is known about the molecular basis of ultradian rhythms. Blum et al. have now identified a second internal clock within the brain, which they name 'the DUO', and shown that this clock normally works in concert with the circadian clock to regulate daily patterns of activity and alertness.

Experiments in mice revealed that the DUO uses the brain chemical dopamine to generate bursts of activity roughly every four hours. Moreover, it continues to work when the circadian clock has been destroyed. Measurements of dopamine in freely moving mice showed that levels of the chemical fluctuate in synchrony with the animals' activity levels. Moreover, drugs that flood the brain with dopamine, such as methamphetamine, disrupt the 4-hour cycle by lengthening the period between bursts of activity, whereas drugs that block dopamine receptors have the opposite effect.

As well as revealing a mechanism by which the brain coordinates processes that repeat several times per day, the identification of the DUO could also provide insights into the biological basis of psychiatric disorders. Conditions such as schizophrenia and bipolar disorder are often accompanied by disturbances in patterns of activity and rest. While these have previously been attributed to the disruption of circadian rhythms, there is little direct evidence for this, which raises the possibility that these changes might instead reflect the disruption of ultradian rhythms.

DOI: 10.7554/eLife.05105.002

(***Gerkema and Verhulst, 1990***). However, despite their prevalence and hypothesized biological significance, ultradian locomotor rhythms have received little research attention, likely owing to their frequently masked expression and unstable periodicity in contrast with circadian activity rhythms (***Ruis et al., 1987***; ***Schibler, 2008***) (***Figure 1C,D***).

In addition to ultradian oscillations, methamphetamine-induced rhythms of locomotor activity occur in the absence of a functional circadian clock. When provided in the drinking water, methamphetamines (Meth) induce a daily activity bout in addition to the expected circadian rhythm (***Honma et al., 1986***). This Meth-dependent component—which typically adopts a period in the circadian range—is not abolished upon SCN lesion nor by genetic disruption of circadian clock function (***Honma et al., 1987***; ***Mohawk et al., 2009***). It was thus concluded that a methamphetamine-sensitive circadian oscillator (MASCO) outside the SCN exists which is capable of driving daily cycles of locomotor activity (***Tataroglu et al., 2006***). Despite the longstanding recognition of ultradian and Meth-dependent rhythms, the underlying cellular and molecular identity of the oscillator(s) driving them is unknown.

Here, we provide evidence for a highly tunable dopaminergic ultradian oscillator (DUO) which is continuously operative in the mammalian brain and which, together with the circadian clock, orchestrates the daily pattern of arousal. Our data suggest that dopamine acts as both the principal oscillator output as well as an integral component of the DUO, determining oscillator period. Our findings further indicate that the previously described MASCO represents a long-period manifestation of the DUO resulting from elevated dopamine tone. Importantly, our data support an intriguing proposition: that DUO, rather than circadian clock, dysregulation critically contributes to the sleep-wake abnormalities associated with psychopathology.

Results

Dopamine transporter deficiency results in ultradian locomotor period lengthening

To gain insights into the mechanistic basis of ultradian locomotor rhythm generation, we considered that locomotor activity is associated with an awakened state (***Welsh et al., 1988***) and consequently,

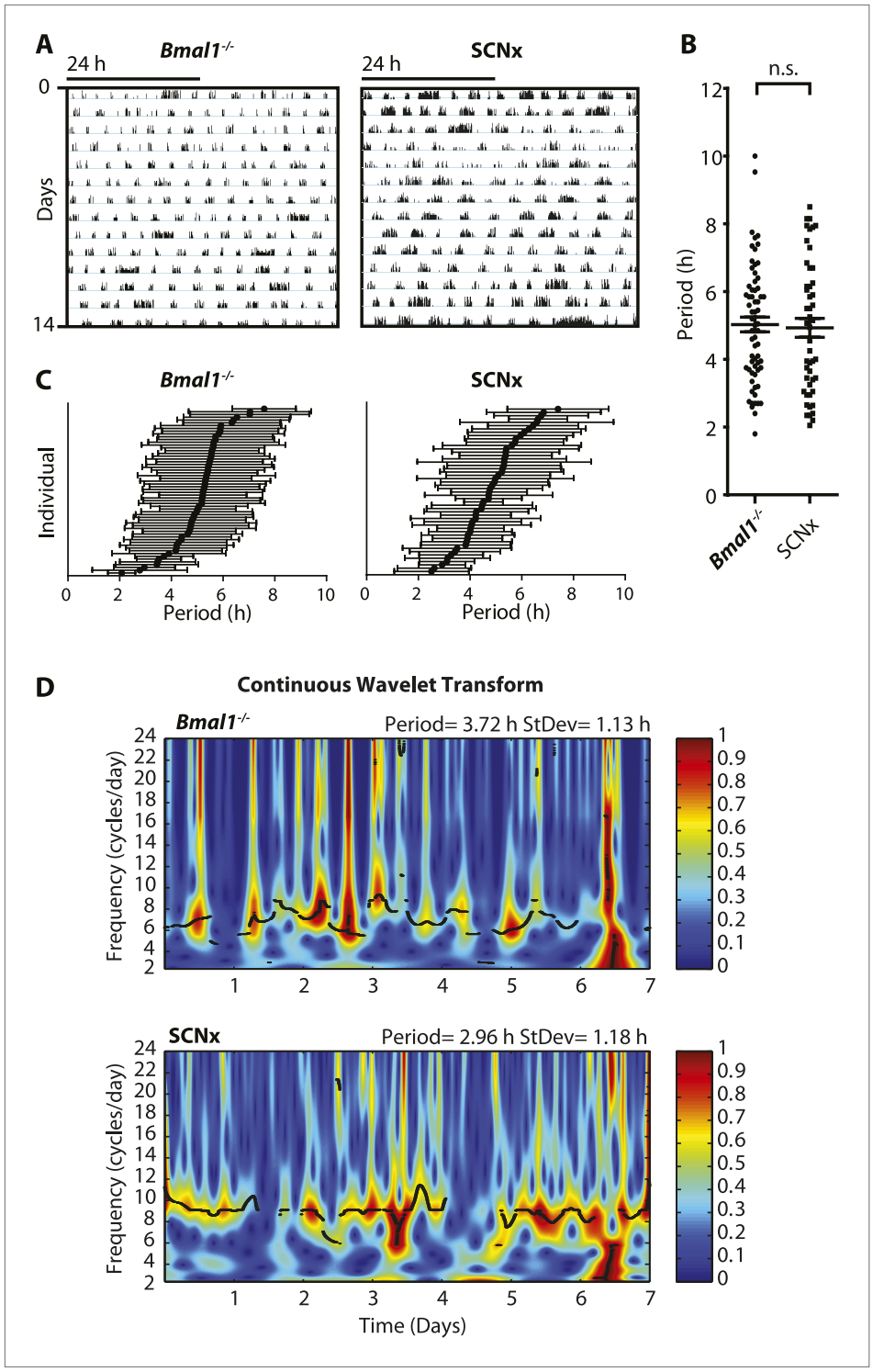


Figure 1. Inter- and intra-animal variability of ultradian activity rhythms across time in circadian incompetent mice. (**A**) Representative double-plotted actograms of running wheel activity in SCN-lesioned (SCNx) and *Bmal1*$^{-/-}$ mice in DD. (**B**) Dot plot of locomotor period length in DD based on Lomb-Scargle periodogram analysis of seven consecutive days of activity recording (N = 65 for *Bmal1*$^{-/-}$ and N = 48 for SCNx; t_{111} = 0.2785, p = 0.78, unpaired t-test). (**C**) Intra-animal period variability expressed as mean ± SD for each animal, ranked according to mean period length derived from continuous wavelet transforms (CWT) for the 1–12-hr frequency range (same animals and timespans for calculation as in **B**). (**D**) CWT-heatmaps showing decibel scaled and normalized amplitude of oscillations according to frequency and time with black traces indicating the ridge of local amplitude maxima.
DOI: 10.7554/eLife.05105.003

the ultradian locomotor rhythms observed in mice that lack circadian clock function (***Figure 1***) could be interpreted as rhythms of heightened wakefulness or arousal. In mammals, a key role in arousal promotion has been attributed to distinct monoaminergic neuronal populations located in the upper brainstem and midbrain (***Brown et al., 2012***). While altering extracellular levels of the arousal-associated monoamines serotonin, norepinephrine, or histamine by genetic manipulation has only limited effects on locomotion (***Thomas and Palmiter, 1997***; ***Bengel et al., 1998***; ***Xu et al., 2000***; ***Parmentier et al., 2002***; ***Zhao et al., 2006***), depleting the brain of dopamine (DA) profoundly abrogates locomotor activity (***Zhou and Palmiter, 1995***). Moreover, increasing extracellular DA levels induces hyperlocomotion (***Giros et al., 1996***) and lengthens the time spent awake (***Wisor et al., 2001***). We therefore speculated that altering DA tone may affect ultradian rhythm generation. To test this, we examined running wheel activity in mice carrying a disruption in the *Slc6a3* gene, which encodes the dopamine transporter (DAT). *Slc6a3*$^{-/-}$ mice exhibit hyperdopaminergia due to the lack of DAT-mediated DA reuptake into dopaminergic neurons, leading to a hyperactivity phenotype (***Giros et al., 1996***; ***Gainetdinov et al., 1998***). As the presence of the circadian clock and/or a light:dark cycle frequently masks ultradian activity rhythms (***Schibler, 2008***), we assessed the locomotor behavioral consequences of DAT elimination in the absence of the master SCN circadian pacemaker. To do so, we electrolytically lesioned the SCN of *Slc6a3*$^{-/-}$ mice and their wildtype littermates and monitored their running wheel behavior in constant darkness (DD). While control mice (SCNx-*Slc6a3*$^{+/+}$) exhibited ultradian activity rhythms with the expected ~4-hr period, SCNx-*Slc6a3*$^{-/-}$ mice showed rhythms whose periods were three times longer (***Figure 2A,B***). Analysis of mice that were deficient for both DAT and the essential clock component BMAL1 (*Bmal1*$^{-/-}$, *Slc6a3*$^{-/-}$) corroborated this finding. *Bmal1*$^{-/-}$, *Slc6a3*$^{-/-}$ mice exhibited ~12–14-hr rhythms in locomotor activity, largely phenocopying the SCNx-*Slc6a3*$^{-/-}$ mice, while their *Bmal1*$^{-/-}$, *Slc6a3*$^{+/+}$ littermates showed ~4-hr periods as expected for isodopaminergic mice lacking circadian clock function (***Figure 2C,D***). Together, these results suggest that DAT removal markedly increases ultradian cycle length. Alternatively, the ~12-hr rhythms observed in SCNx-*Slc6a3*$^{-/-}$ and *Bmal1*$^{-/-}$, *Slc6a3*$^{-/-}$ animals may originate from an independent oscillator, one that is activated by DAT elimination, while the short period ultradian oscillator that operates in DAT intact, SCNx *or Bmal1*$^{-/-}$ animals is disengaged or otherwise obscured.

The psychostimulant methamphetamine lengthens ultradian locomotor period

In order to corroborate that DAT removal indeed lengthens the period of the ultradian activity cycles, we took into account that DAT-mediated DA uptake can be reversed by the selective action of the psychostimulant methamphetamine (Meth) (***Howell and Kimmel, 2008***). Because Meth leads to increased extracellular DA concentrations as in the case of *Slc6a3* gene disruption, we speculated that Meth treatment would result in a, possibly gradual, period lengthening of the ultradian locomotor rhythms. Indeed when we treated *Bmal1*$^{-/-}$ animals with increasing concentrations of Meth via drinking water in DD, we observed a gradual lengthening of the initial ~4-hr locomotor oscillations and this was accompanied by a corresponding increase in activity bout length (***Figure 3A,B***). Of note, the period increase in response to elevating Meth concentrations did not halt in the circadian range, rather, oscillations continued to lengthen with individual animals reaching periods of 100 hr or more (***Figure 3C***). Gradual period lengthening of ultradian locomotor rhythms was also observed in *Bmal1*$^{-/-}$ animals exposed to amphetamine (***Figure 3—figure supplement 1A,B***), a drug similarly targeting DAT (***Howell and Kimmel, 2008***), but with lower efficacy than Meth (***Goodwin et al., 2009***). These results argue that DAT targeting psychostimulants affect an endogenous ultradian rhythm generator by increasing period length. However, due to the mode of delivery (drinking water) and the rhythmic Meth uptake that may in turn result from it, it is conceivable that the Meth-dependent, long-period oscillations are 'driven' by rhythmic drug intake rather than being generated endogenously. To address this possibility, we subcutaneously implanted *Bmal1*$^{-/-}$ mice with osmotic minipumps that continuously infused Meth over a period of 2 weeks. Running wheel analysis in DD of Meth-infused *Bmal1*$^{-/-}$ animals demonstrated a significant ultradian period lengthening upon drug infusion (***Figure 3—figure supplement 1C,D***) suggesting that rhythmic uptake is not required for Meth to exert its period lengthening effect, in line with a previous study performed in rats (***Honma et al., 1987***). The relatively limited change in period observed in this experimental paradigm could be due to the short, 2-week infusion timespan, which is perhaps insufficient to robustly lengthen periods beyond 12 hr (***Figure 3—figure supplement 1D***). Together, these findings support the notion that the long-period oscillations observed

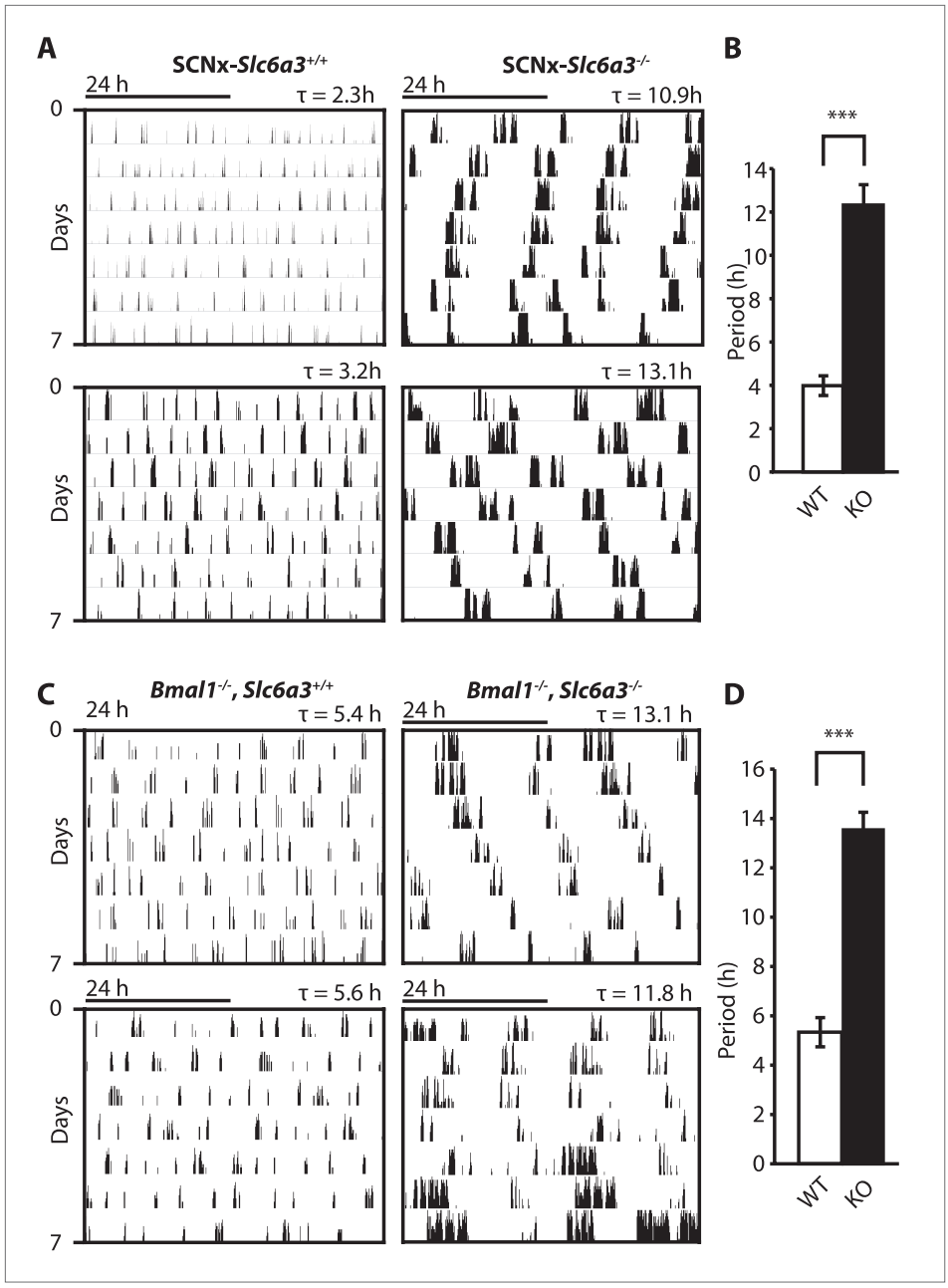


Figure 2. Dopamine transporter knockout alters periodicity of ultradian locomotor rhythms in circadian incompetent mice under constant darkness. (**A**) Representative, double-plotted actograms demonstrating marked lengthening of ultradian locomotor periods in SCNx-*Slc6a3*$^{-/-}$ mice as compared to SCNx-*Slc6a3*$^{+/+}$ littermates. Tau(τ) indicates individual period. (**B**) Period length averages of ultradian activity (N = 7; $F_{1,6}$ = 253.8, ***p < 0.0001, ANOVA) from Lomb-Scargle periodogram analysis of 7-day time-spans. (**C**) Representative, double-plotted actograms demonstrating markedly increased ultradian period lengths in *Bmal1*$^{-/-}$, *Slc6a3*$^{-/-}$ mice as compared to *Bmal1*$^{-/-}$, *Slc6a3*$^{+/+}$ mice. (**D**) Period length averages of ultradian activity (N = 4; $F_{1,3}$ = 194.2, ***p < 0.001, ANOVA) from Lomb-Scargle periodogram analysis of 7-day time- spans.

DOI: 10.7554/eLife.05105.004

in Meth-treated animals, and likewise in *Slc6a3*$^{-/-}$ mice, are due to period expansion of an endogenous ultradian rhythm of arousal.

We next aimed to confirm these findings in intact mice. While activity rhythms in the ultradian range are easily discernable in voles and hamsters (***Gerkema et al., 1990***; ***van der Veen et al., 2006***; ***Prendergast et al., 2012***), the presence of such rhythms in mice or rats is much less obvious and this

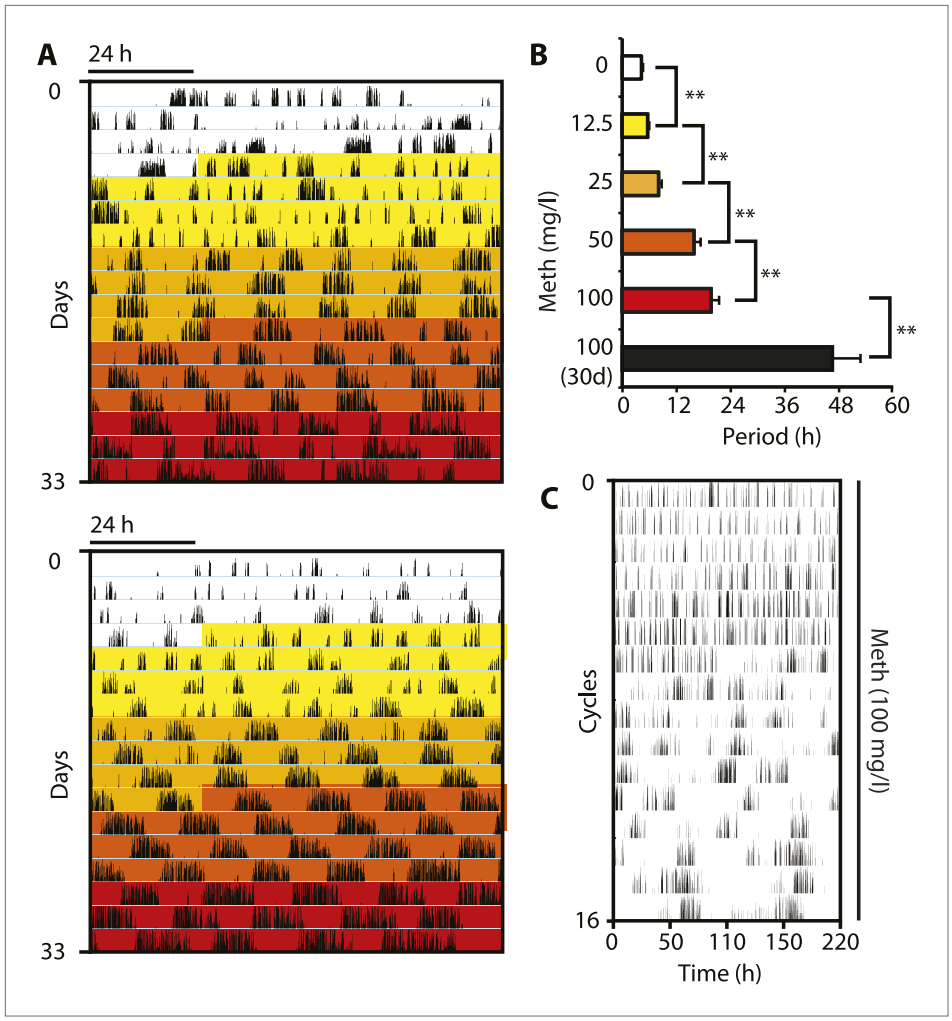


Figure 3. Pharmacological tuning of DAT activity by incrementally increasing methamphetamine concentrations lengthens ultradian locomotor period into the infradian range. (**A**) Representative actograms of Meth-treated *Bmal1*$^{-/-}$ mice in DD. Treatment intervals are highlighted with corresponding concentrations indicated in (**B**). (**B**) Mean periods from the last 7 days at a given Meth concentration. Repeated measures ANOVA revealed a significant main effect of concentration/time ($F_{5,40}$ = 34.30, $p < 0.001$) and significant period lengthening between consecutive concentrations (mean ± SEM; N = 9; **$p \leq 0.005$, planned comparisons; 30 day, 30 day exposure to 100 mg/l). (**C**) Modulo 110-hr actogram of a *Bmal1*$^{-/-}$ animal after extended exposure to Meth revealing an ultra-long activity rhythm.

DOI: 10.7554/eLife.05105.005

The following figure supplement is available for figure 3:

Figure supplement 1. Amphetamine treatment lengthens ultradian locomotor rhythms in *Bmal1*$^{-/-}$ mice.

DOI: 10.7554/eLife.05105.006

has been attributed to masking by the circadian clock (***Schibler, 2008***). Indeed, running wheel activity data from intact mice often do not provide strong indications of ultradian activity rhythms. However, when we recorded ambulatory behavior using telemetry, we frequently detected three, evenly spaced activity bouts during the active (night) phase (***Figure 4A***). The observed ~4 hr peak-to-peak spacing is in line with a previous study on several circadian competent mouse strains where ultradian rhythms with similar period length were detected (***Dowse et al., 2010***). The absence of a clearly discernable ~4-hr rhythmic component during the light portion of the daily LD cycle (***Figure 4A***) is likely due to masking by the SCN and/or light. Unlike WT mice, *Slc6a3*$^{-/-}$ mice never showed a triple-peak activity pattern at night (***Figure 4B***), and spectral density analysis in the ultradian (2–8 hr) range (***Figure 4C***) underscores that rhythm generation is significantly altered in these animals while total daily activity is not (***Figure 4D***). When we provided C57BL/6 mice carrying a telemetry implant with Meth in their

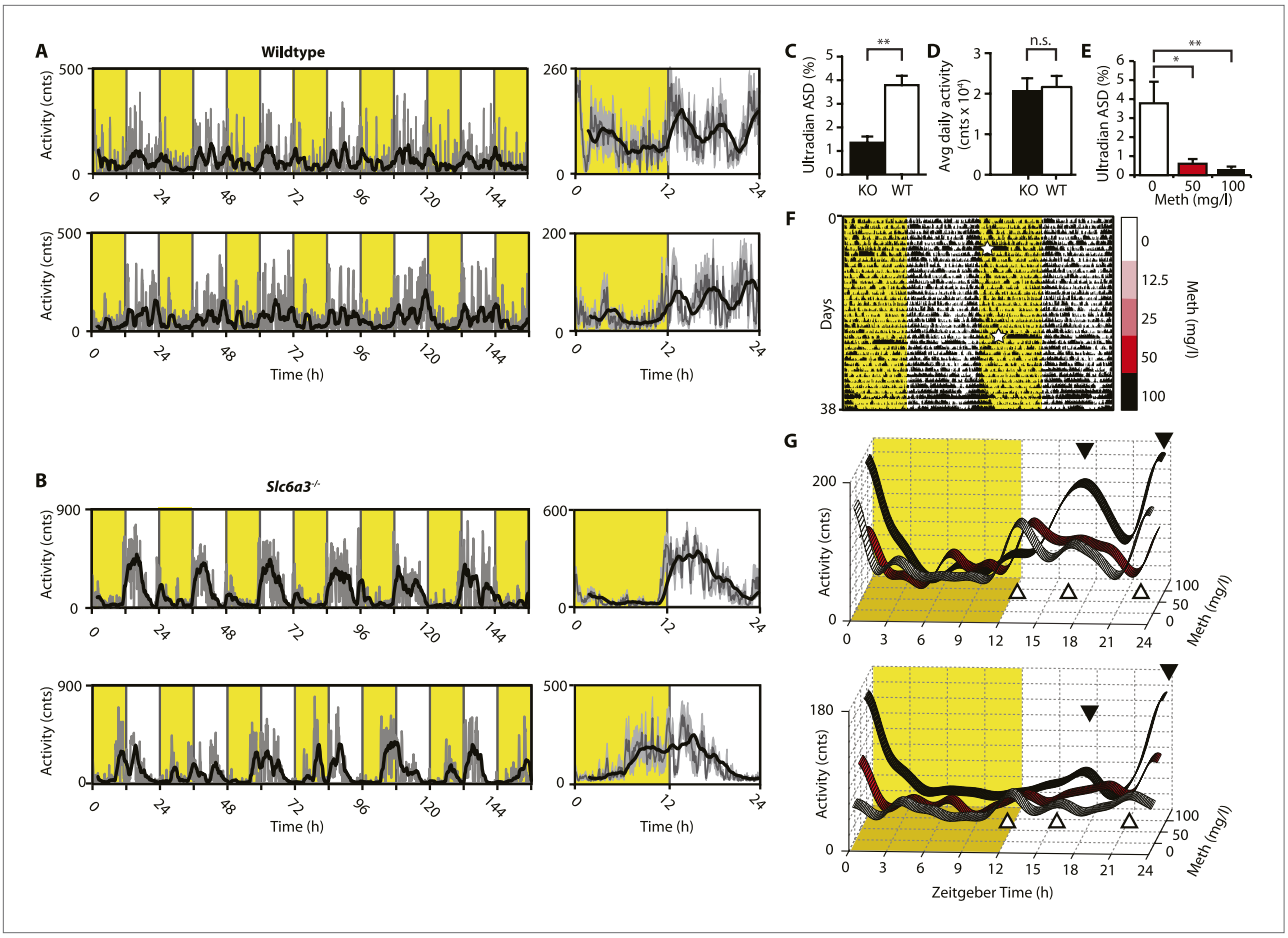


Figure 4. Ultradian activity in SCN-intact *Slc6a3*$^{-/-}$ mice and their wildtype littermates. (**A** and **B**) Ambulatory activity recorded by telemetry implants across multiple days (left) and averaged over 24 hr (right). Traces represent 2-hr recursive smoothing (black) of the underlying raw DATa (dark grey; SEM envelope, light grey). Areas in yellow indicate lights on. (**C**) Amplitude spectral density in the ultradian range (2–8 hr) is significantly different between *Slc6a3*$^{-/-}$ mice and their wildtype littermates revealing a loss of the ultradian component (mean ± SEM; N = 10; $F_{1,18}$ = 26.40, **p < 0.0001, ANOVA). (**D**) Although the temporal pattern of locomotion differs between genotypes there is no significant difference in daily activity averaged over multiple days (mean ± SEM; N = 10, $F_{1,18}$ = 0.1793, ANOVA). (**E**–**G**) Addition of Meth to the drinking water of C57BL/6 mice lengthens the night-time activity bouts in a concentration-dependent manner. Averaged daily locomotor activity of individual mice at different Meth concentrations (**G**) derived from the time-span indicated by colored bars next to the representative actogram (**F**) and subjected to Butterworth filtering. The three night-time activity peaks before treatment (white triangles), transform to 2 peaks after exposure to the highest concentration (black triangles). The night-time bout lengthening is also reflected in the reduction of ultradian amplitude spectral density in the 1 to 5-hr range (**E**, mean ± SEM, N = 7; $F_{2,20}$ = 8.08, p < 0.005, repeated measures ANOVA; *p < 0.05, **p < 0.01, post-hoc Bonferroni). White asterisks (in **E**) indicate cage changes.

DOI: 10.7554/eLife.05105.007

drinking water, we observed a gradual lengthening of the interval between the night-time activity peaks (***Figure 4F,G***), resulting in the transformation of the triple peak (***Figure 4G***, white trace and white triangles) into a dual or single night-time activity peak pattern at the highest (100 mg/l) Meth concentration (***Figure 4G***, black trace and black triangles). This was corroborated by spectral density analysis revealing a marked reduction in ultradian frequencies upon Meth treatment (***Figure 4E***). Together, these results indicate that blocking DA reuptake also lengthens ultradian activity bout intervals in circadian competent mice, arguing that an ultradian oscillator is not 'activated' by circadian disruption, but rather continuously operative alongside the circadian timer. As with *Bmal1*$^{-/-}$ mice, prolonged Meth-treatment in intact WT mice can lead to profound period lengthening of ultradian rhythms, often into the circabidian (48 hr) range (for examples see Figures 5E, 7F, 9B).

The antipsychotic haloperidol shortens ultradian locomotor period

Given that both DAT disruption and Meth treatment elevate extracellular DA and concurrently lengthen ultradian rhythm period, we speculated that manipulations aimed at lowering DA tone

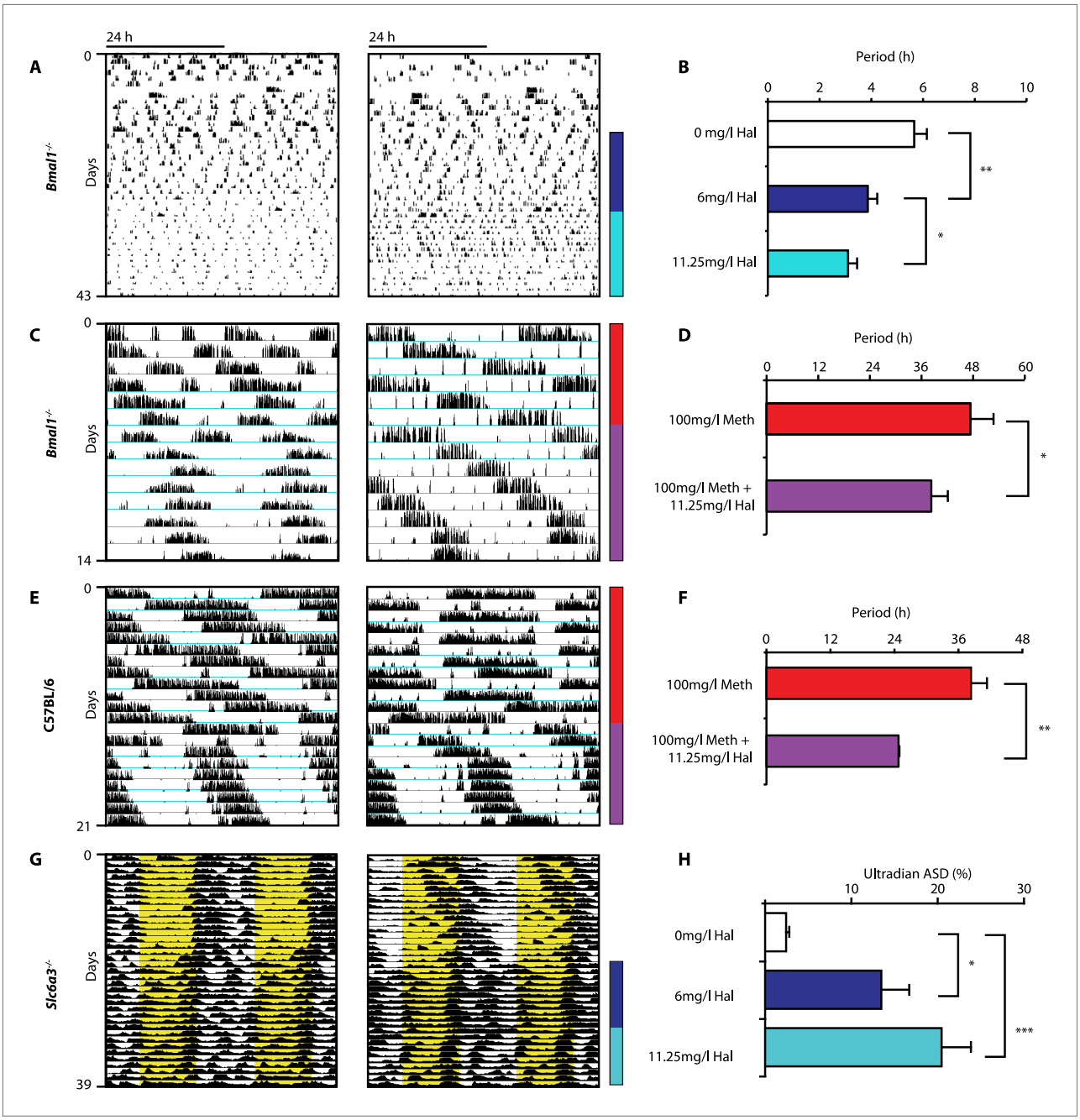


Figure 5. Haloperidol shortens circadian-clock-independent locomotor rhythms. Representative actograms with Hal treatment periods indicated by colored bars (left); bar graphs indicate corresponding locomotor period (right). (**A**) Increasing concentrations of Hal provided in the drinking water gradually shortens the endogenous ultradian activity rhythms of *Bmal1*$^{-/-}$ mice in DD (**B**, mean ± SEM, N = 12; $F_{2,10}$ = 14.36, p < 0.0001, repeated measures ANOVA; *p < 0.05 and **p < 0.01, planned comparison ANOVA). (**C** and **E**) Hal shortens the infradian rhythms in Meth-treated *Bmal1*$^{-/-}$ mice (**D**, mean ± SEM, N = 9; $F_{1,8}$ = 2.357, *p < 0.05 ANOVA) and WT mice (**F**, mean ± SEM, N = 9; $F_{1,8}$ = 3.525, **p < 0.01, ANOVA) in DD. (**G**) Hal treatment increases the frequency of temperature fluctuation in *Slc6a3*$^{-/-}$ mice under LD measured by telemetric implants. (**H**) Changes of amplitude spectral density in the ultradian range (2–8 hr) in response to increasing Hal concentrations (mean ± SEM; N = 8; $F_{2,7}$ = 14.74, repeated measures ANOVA, *p < 0.05, ***p < 0.0005, post-hoc Bonferroni). For all experiments, periods are calculated based on the running wheel activity during the last 7 days of treatment at the indicated Meth/Hal concentration.

DOI: 10.7554/eLife.05105.008

should conversely lead to period shortening. We therefore provided *Bmal1*$^{-/-}$ mice in DD with the antipsychotic drug haloperidol (Hal) in their drinking water. Hal selectively blocks the DA receptor D2 (DRD2) (***Strange, 2008***), which is expressed on postsynaptic target sites of DA neuronal projections

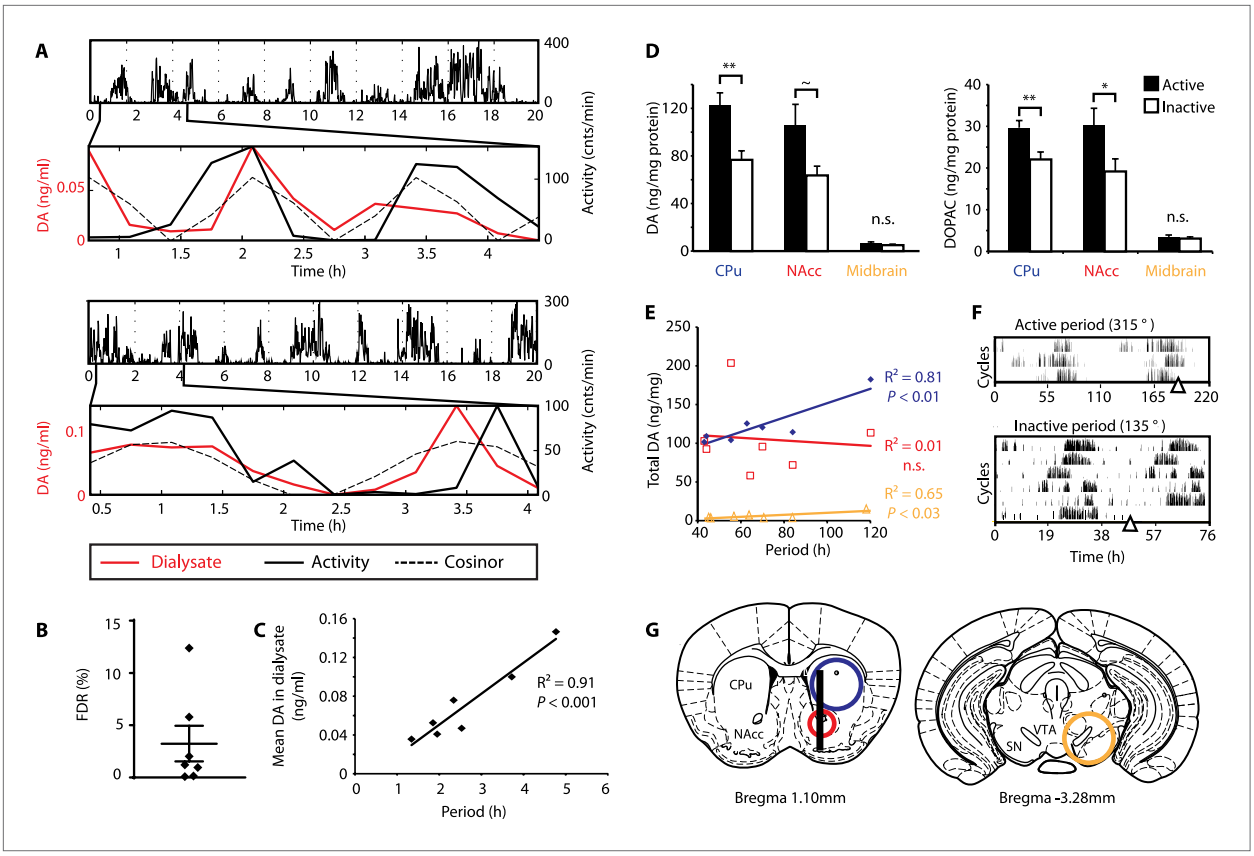


Figure 6. DA fluctuations correlate with ultradian locomotor behavior in *Bmal1*$^{-/-}$ mice. (**A**) Two representative examples of in vivo striatal microdialysis in *Bmal1*$^{-/-}$ mice. Upper: locomotor activity as measured by beam breaks. Lower: DA dialysate concentration measured at 20-min intervals (red trace) plotted alongside the corresponding locomotor activity (solid black trace). Cosinor (dotted line) was computed from the 20 hr of locomotor activity following dialysate sampling. (**B**) False discovery rate of the fit between the DA profiles and corresponding cosinors (mean ± SEM, N = 7). (**C**) Linear regression analysis of period length vs mean DA concentration, dots representing individual animals. (**D**) Tissue punches of CPu, NAcc, and midbrain were analyzed for DA and DOPAC content in animals sacrificed during their active (■) vs their non-active (□) phase (mean ± SEM; active phase, N = 7, inactive phase, N = 6; *p < 0.05, **p < 0.01, ~p = 0.066, ANOVA). (**E**) Linear regression for period vs DA content in animals sacrificed during the active phase revealed significant correlations for the CPu and midbrain. (**F**) Representative actograms displaying activity rhythms of Meth-treated *Bmal1*$^{-/-}$ mice used for tissue collection. Triangles indicate collection time points (in °), with locomotor activity bout onset set to 180°. (**G**) Illustrations of coronal mouse brain sections (***Paxinos and Franklin, 2001***) indicating position of the active membrane (black bar) and tissue punch placements (circles, colors correspond to labels in **D** and **E**).

DOI: 10.7554/eLife.05105.012

The following figure supplement is available for figure 6:

Figure supplement 1. Rhythms of extracellular DA in the striatum of freely-moving *Bmal1*$^{-/-}$ mice.

DOI: 10.7554/eLife.05105.013

but also presynaptically as an autoreceptor (***Schmitz et al., 2003***). Importantly, when given chronically as in our case, Hal has been reported to electrically silence midbrain DA neurons (***White and Wang, 1983***) and to markedly lower extracellular DA levels in the striatum/nucleus accumbens regions in rats (***Lane and Blaha, 1987***; ***Ichikawa and Meltzer, 1991***). As predicted, *Bmal1*$^{-/-}$ mice responded with successive locomotor period shortening to increasing concentrations of Hal (***Figure 5A,B***). We also observed Hal-mediated shortening of the long-period behavioral rhythms in *Bmal1*$^{-/-}$ and wild-type mice treated concurrently with Meth (***Figure 5C–F***). A period shortening effect of Hal could also be discerned from the core body temperature fluctuations of *Slc6a3*$^{-/-}$ mice, which increased in frequency (***Figure 5G***). This response, which mirrored the locomotor behavioral response (not shown), was concentration dependent and spectral density analysis confirmed a successive increase of the ultradian component upon Hal exposure (***Figure 5H***). Together, these data strongly support the hypothesis that both the ultradian rhythms observed in wildtype and *Bmal1*$^{-/-}$ mice and the long-period rhythms previously attributed to the MASCO are driven by the exact same oscillatory mechanism.

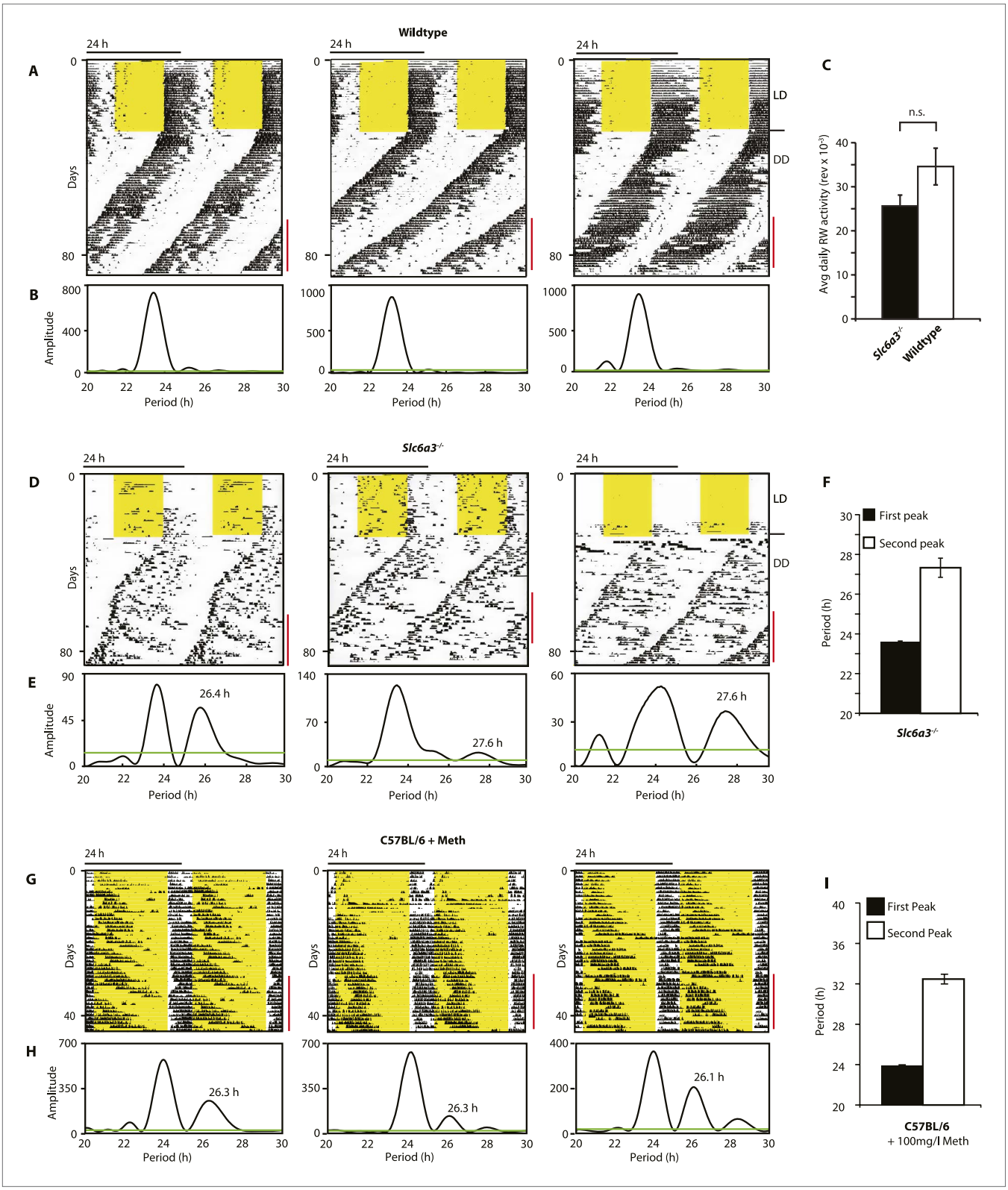


Figure 7. *Slc6a3*$^{-/-}$ mice show a second rhythmic locomotor activity component. Representative actograms displaying daily running wheel activity of *Slc6a3*$^{+/+}$ (**A**), *Slc6a3*$^{-/-}$ mice (**D**) and C57BL/6 mice on Meth (**G**). (**B**, **E** and **H**) Lomb-Scargle periodograms generated from the time-span indicated by red bars, in (**A**, **C**, **G**), respectively. (**C**) There is no significant difference between genotypes in daily activity averaged over the time-span of analysis (mean ± SEM; N = 6, $F_{1,10}$ = 1.848, ANOVA) (**F** and **I**) Average periods of the highest 2 periodogram peaks for *Slc6a3*$^{-/-}$ mice (**E**, mean ± SEM, N = 6)

Figure 7. Continued on next page

Figure 7. Continued

and C57BL/6 mice on Meth (**H**, mean ± SEM, N = 9, Meth-treatment started on day 1 of the recordings). Areas in yellow indicate lights-on. Green line in the periodograms indicates the confidence threshold for rhythmicity ($\alpha = 0.01$).

DOI: 10.7554/eLife.05105.009

The following figure supplement is available for figure 7:

Figure supplement 1. The SCN of *Slc6a3*$^{-/-}$ mice is unperturbed.

DOI: 10.7554/eLife.05105.010

As altering extracellular DA is a common denominator of all three manipulations—DAT elimination, (meth) amphetamine, and haloperidol treatment—these findings collectively suggest that DA tone determines ultradian period. Given that DA is known to mediate arousal/wakefulness (***Brown et al., 2012***), it appears plausible that DA also serves as principal oscillator output, driving rhythms in arousal. We thus hypothesized that extracellular DA must oscillate in synchrony with the observed activity rhythms.

Extracellular DA fluctuates in synchrony with ultradian activity cycles

To test this hypothesis, *Bmal1*$^{-/-}$ mice were unilaterally implanted with a microdialysis probe positioned along the ventro-dorsal extent of the right striatum (***Figure 6G***), a site heavily innervated by DA neurons. Mice were kept under constant dim red light (<5lux) and ultradian activity rhythms were monitored using infra-red beam breaking. The animals showed the expected short period rhythms in locomotor activity throughout the experiment (***Figure 6A***), and analysis of dialysates revealed that extracellular DA levels fluctuated concordantly with the activity cycles (***Figure 6A*** and ***Figure 6—figure supplement 1***, compare solid red and black traces). Importantly, we generated a model oscillation using the cosinor method from locomotor rhythms recorded for 20 hr after dialysate sampling with no investigator present and compared it to the DA fluctuations observed. The goodness of fit between the cosinor and the DA profiles indicated by the sum of squared errors (SSE = 0.050 ± 0.027, mean ± SEM, N = 7) suggests that the dialysate sampling procedure itself did not perturb the generation of the endogenous ultradian activity cycles. To determine the statistical significance of the observed agreement between the cosinor and the DA trace, we randomly permutated the individual time points of the DA trace and determined the percentage of permutated traces with an equal or better fit to the cosinor in comparison to the observed DA fluctuation. On average, only 3.2 ± 1.7% of 100,000 permutated DA profiles correlated as well or better than the experimentally observed DA measurements (***Figure 6B***), confirming that extracellular DA fluctuates in synchrony with the ultradian activity cycles. Of particular note, linear regression revealed a highly significant correlation between mean DA concentration in the dialysate and ultradian locomotor period in the *Bmal1*$^{-/-}$ animals we tested (***Figure 6C***), which again supports a role of (extracellular) DA as a period determinant of the ultradian activity cycle.

As our findings suggest that the observed >24-hr activity rhythms are due to period lengthening of the ultradian oscillations, we similarly expected DA levels to fluctuate in synchrony with the activity cycles in Meth-treated animals. To test this, we measured DA content in tissue punches of Meth-treated *Bmal1*$^{-/-}$ mice from the dorsal (caudate putamen, CPu) and ventral (nucleus accumbens, NAcc) striatum as well as the ventral midbrain, which includes both the substantia nigra (SN) and the ventral tegmental area (VTA) (***Figure 6G***). For punch collection, animals were sacrificed during their active and inactive phases, respectively (***Figure 6F***). We detected significantly higher levels of DA as well as its immediate metabolite, 3,4-dihydroxyphenylacetic acid (DOPAC), in tissue extracts of the CPu during the animal's active phase (***Figure 6D***). There was also a significant increase in DOPAC levels and a trend towards elevated DA (p = 0.066) during the active period in the NAcc, while extracts from the midbrain region, which contains the cell bodies of the striatal dopaminergic afferents, exhibited no detectable change (***Figure 6D***).

We also found a significant correlation between locomotor period and DA content in CPu and midbrain punches collected during the active phase (***Figure 6E***), again consistent with a role for DA as a period determinant of ultradian oscillations.

SCN-intact *Slc6a3*$^{-/-}$ mice show two components of rhythmic activity

As *Slc6a3*$^{-/-}$ mice lack the triple peak night-time activity pattern that is characteristic of wild-type mice, we subjected *Slc6a3*$^{-/-}$ mice to long-term running wheel activity monitoring to examine their

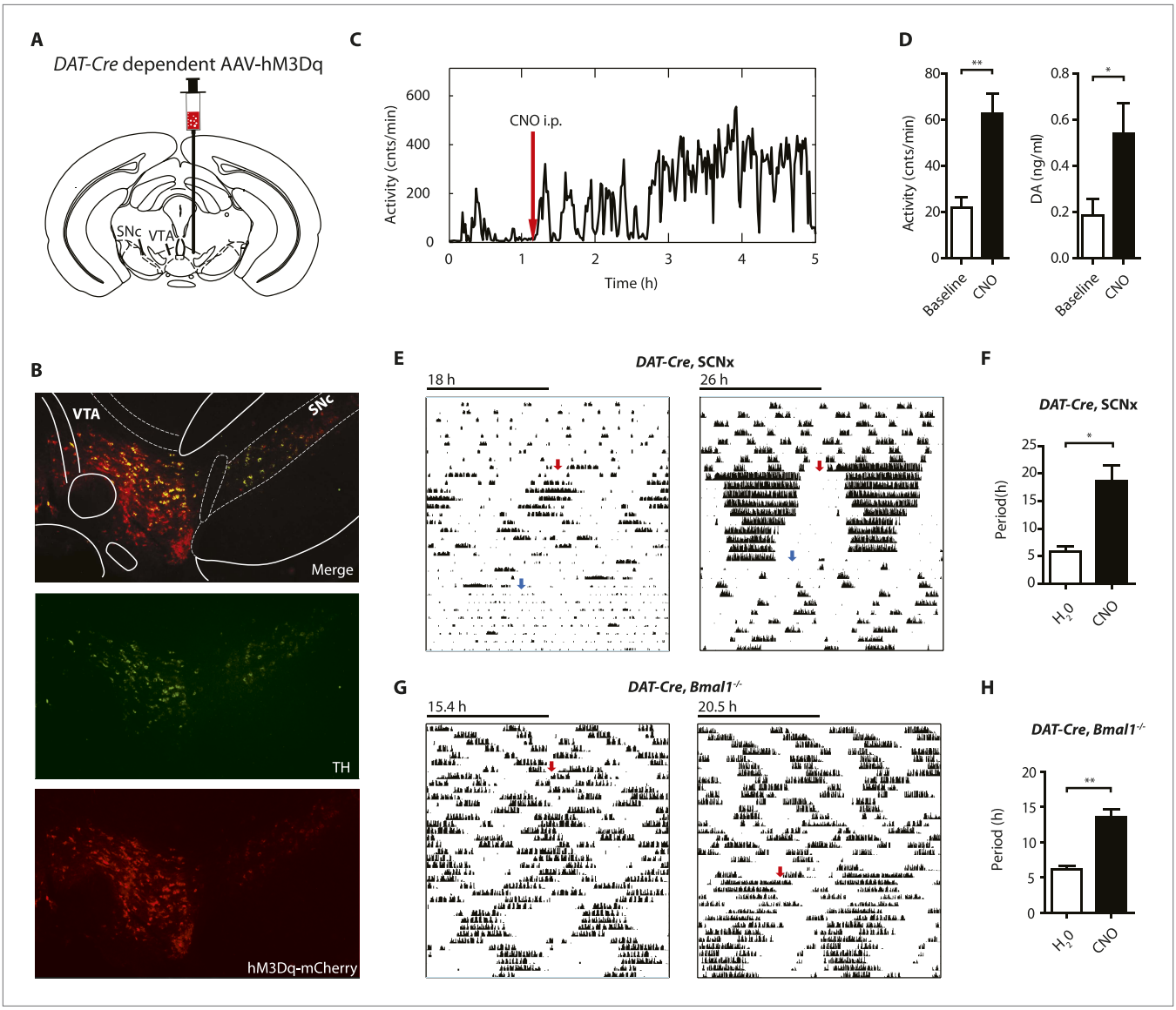


Figure 8. Chemogenetic activation of midbrain DA neurons lengthens ultradian locomotor period. (**A**) *AAV-DIO-hM3Dq-mCherry* was stereotaxically and bilaterally delivered into the VTA/SN region of *DAT-Cre* transgenic mice as indicated. (**B**) Representative immuno-fluorescent image of the ventral midbrain from a virus-injected and behaviorally responsive mouse showing extensive co-expression of the mCherry fusion-tag in TH-positive cells of the midbrain. (**C**) Locomotor response to CNO (red arrow, 1 mg/kg body weight i.p.) of a representative, *AAV-hM3Dq* transduced *DAT-Cre* mouse undergoing microdialysis. (**D**) 20min- binned locomotor activity and extracellular striatal DA content (20 min) of *AAV-hM3Dq* transduced mice 1 hr prior (Baseline) and 2 hr after CNO injection. Mice were implanted with a striatal microdialysis probe and DA content was measured as in ***Figure 6A*** (mean ± SEM; N = 3, *p < 0.05, **p < 0.01, paired t-test). (**E**–**H**) Representative actograms of *AAV-hM3Dq* transduced *DAT-Cre*, SCNx (**E**) and *DAT-Cre*, *Bmal1*$^{-/-}$ mice (**G**). Switch to CNO-containing drinking water (7.5 mg/l) is indicated by red arrow. Blue arrow marks return to regular water (in **E**). Animal activity is plotted modulo according to the period measured during the last 7 days of CNO treatment. Periodogram analysis reveals CNO-dependent period lengthening in both *DAT-Cre*, SCNx (**F**, mean ± SEM; N = 6, $F_{1,5}$ = 3.68,*p < 0.05, ANOVA) and *DAT-Cre*, *Bmal1*$^{-/-}$ (**H**, mean ± SEM; N = 11; $F_{1,10}$ = 20.48, **p < 0.0001, ANOVA) mice. VTA, ventral tegmental area; SNc, substantia nigra pars compacta.

DOI: 10.7554/eLife.05105.014

The following figure supplement is available for figure 8:

Figure supplement 1. *AAV-h3MDq* targeting of DA neurons.

DOI: 10.7554/eLife.05105.015

locomotor behavior in more detail. Under LD conditions, we observed more fragmented daily activity in *Slc6a3*$^{-/-}$ than their wildtype counterpars (***Figure 7A,D***). Upon release into constant darkness (DD), wild-type animals exhibited a daily rhythm with a period slightly shorter than 24 hr as expected for mice (***Bunger et al., 2000***) (***Figure 7A***). While this principal locomotor component was also observed

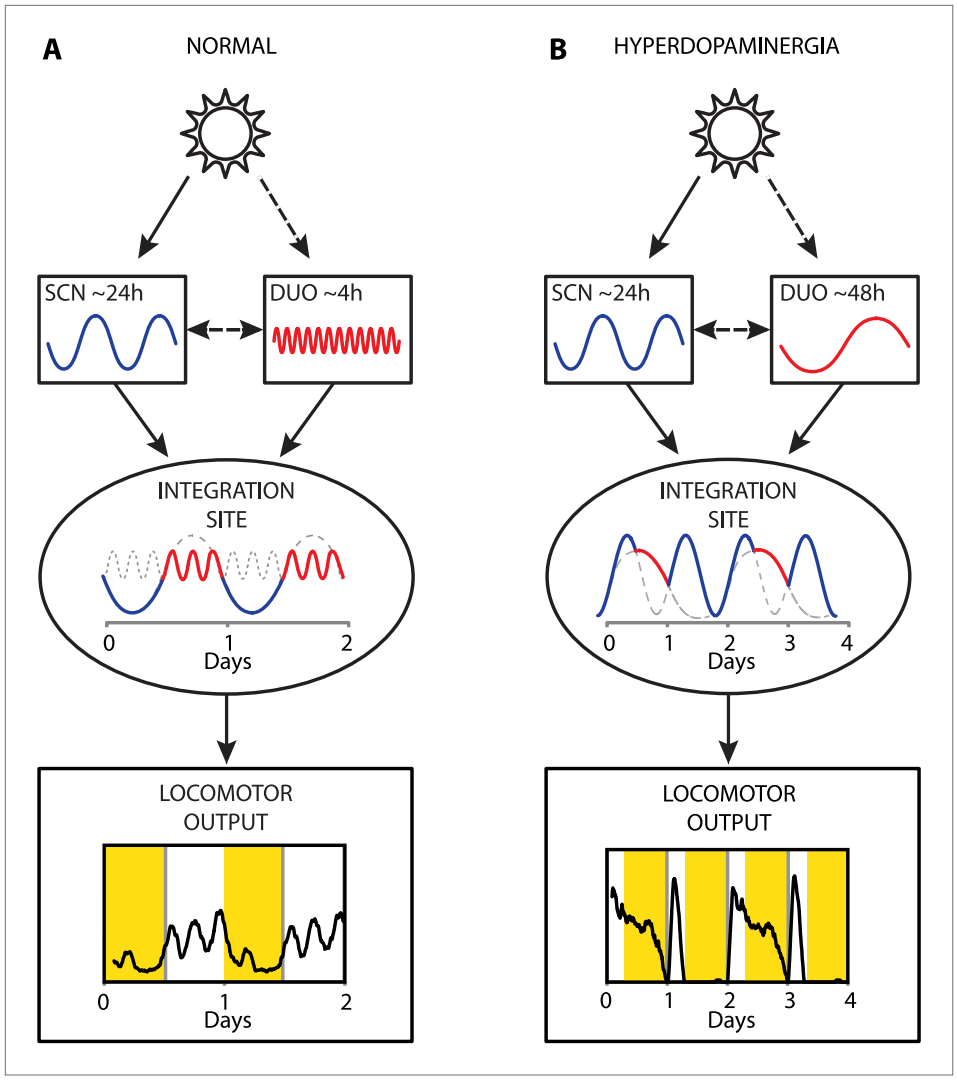


Figure 9. Proposed model of circadian and ultradian oscillator output integration to govern daily locomotor behavior. (**A**) Light input entrains the circadian (SCN) and (indirectly) the ultradian (DUO) oscillators creating a stable phase relationship. Their rhythmic outputs, upon integration at a common downstream effector, generate the daily pattern of locomotor activity. (**B**) Under conditions of high DA tone (e.g., DAT elimination or Meth-treatment), DUO period lengthens, leading to a second, separate activity rhythm. This rhythm either free-runs (see ***Figure 7D,G***) or phase locks with the SCN pacemaker by adopting a subharmonic, that is, 48-hr period, as frequently observed upon Meth treatment. Representative output plots show average activity of individual mice computed from 8 days of ambulatory activity (**A**) or 14 days of running wheel activity under Meth treatment (**B**).
DOI: 10.7554/eLife.05105.011

in *Slc6a3*$^{-/-}$ mice, they repeatedly exhibited an additional activity component with periods longer than 24 hr (***Figure 7D–F***). This component, which was more evident after the first few weeks in DD, extended from the main activity component and lasted for several cycles before disappearing late in the subjective day (***Figure 7D***). Overall, there was no significant difference in the total amount of daily activity even though the temporal pattern of locomotion in these mice is severely altered (***Figure 7C***). The observed dual component activity pattern shows striking similarities to the activity patterns of Meth-treated C57BL/6 mice (***Figure 7G–I***), suggesting that the second rhythmic component is generated by the dopaminergic oscillator.

However, such locomotor activity pattern, with two components oscillating at different frequencies, has also been observed in rats housed under a 22-hr LD cycle and have been attributed to a dissociation of molecular rhythms between the dorsal and ventral SCN (***de la Iglesia et al., 2004***). To rule out that the SCN is similarly affected in *Slc6a3*$^{-/-}$ mice, we sacrificed mice at the peak and trough of the

main activity component when the second rhythm was maximally antiphasic (***Figure 7—figure supplement 1A***). In situ hybridization directed against *Period1* transcripts did not indicate rhythmic desynchrony between or within SCN hemispheres. The riboprobe showed strong, homogenous staining throughout the SCN in the early subjective day, whereas little to no signal was detectable in the early subjective night, suggesting that SCN pacemaker function is unperturbed in *Slc6a3*$^{-/-}$ mice (***Figure 7—figure supplement 1B***).

Together, these findings thus argue that the second rhythmic component of *Slc6a3*$^{-/-}$ mice is indeed a product of the above described dopaminergic oscillator.

Chemogenetic activation of DA neurons lengthens ultradian period

Since striatal and midbrain DA levels vary concordantly with activity rhythms in *Bmal1*$^{-/-}$ mice, and DA neurons of the SN/VTA are by and large the exclusive source of striatal and midbrain DA, these neurons are a likely site of ultradian rhythm regulation, if not generation. To corroborate the importance of these neurons in the oscillator process, we employed a chemogenetic approach based upon DREADD (designer receptors exclusively activated by designer drugs) technology (***Krashes et al., 2011***). We stereotaxically delivered the adeno-associated virus *AAV-DIO-hM3Dq-mCherry* (***Krashes et al., 2011***) into the SN/VTA region of mice that carried an *Slc6a3* promoter driven *Cre* transgene (*DAT-Cre*) (***Figure 8A***) (***Turiault et al., 2007***) and that additionally were either BMAL1-deficient (*DAT-Cre*, *Bmal1*$^{-/-}$) or received an SCN lesion (*DAT-Cre*, SCNx). Upon Cre-mediated recombination, virally transfected cells express the stimulatory DREADD hM3Dq, which has been demonstrated to increase neuronal firing frequency upon binding of the compound clozapine-N-oxide (CNO), which is otherwise physiologically inert (***Alexander et al., 2009***). As augmenting firing frequency in DA neurons enhances DA release (***Sulzer, 2011***), and CNO stimulation of successfully transduced *DAT-Cre* mice evokes vigorous firing of dopamine neurons (***Wang et al., 2013***), CNO is expected to increase DA release. Consistently, i.p.-injection of CNO elevated extracellular DA in the striatum of those mice and led to a prolonged increase in locomotor activity (***Figure 8C,D***), which was absent in vehicle injected animals (***Figure 8—figure supplement 1D***). Responsive mice showed mCherry fusion tag expression selectively in tyrosine hydroxylase (TH) positive cells (***Figure 8B***, ***Figure 8—figure supplement 1A***). Cell counting revealed that, 92.9 ± 5.1% and 87.8 ± 6.0% (mean ± SEM, N = 6 for both groups) of the midbrain TH^{+} neurons also expressed the mCherry fusion tag in responsive *DAT-Cre*, SCNX and *DAT-Cre*, *Bmal*$^{-/-}$ mice, respectively, confirming effective Cre-mediated hM3Dq expression in midbrain DA neurons. Running wheel activity monitoring of virus-injected mice showed the expected short period locomotor oscillations, however, upon switching to CNO-containing drinking water (red arrows), both, *AAV-hM3Dq* transduced *DAT-Cre*, SCNx and *DAT-Cre*, *Bmal1*$^{-/-}$ mice responded with locomotor period lengthening (***Figure 8E–H***), an effect that was reversed when mice were returned to pure water (***Figure 8G***, blue arrows). As expected, a period lengthening was not observable in *AAV-hM3Dq* injected *DAT-Cre*$^{0/0}$, *Bmal1*$^{-/-}$ mice upon CNO exposure (***Figure 8—figure supplement 1B,C***). These results further corroborate that elevating DA tone, in this case by selective excitation of DA neurons, lengthens ultradian locomotor period and that the oscillator driving ultradian rhythmicity comprises DA neurons of the VTA/SN region.

Discussion

Collectively, our results provide strong evidence that a dopaminergic ultradian oscillator (DUO) driving rhythms of behavioral arousal is continuously operative in the mammalian brain. We propose that under normal conditions, this DUO cycles in harmony with the circadian SCN pacemaker and that the rhythmic information of both the SCN and the DUO are integrated at a common downstream site to create the daily pattern in locomotor activity (***Figure 9A***). However, elevation of DA tone can lead to DUO period lengthening, which either results in DUO free-run (for example, ***Figure 7D,G***) or reinstatement of oscillator synchrony albeit at a different harmonic (***Figure 9B***, e.g., 48 hr). The DUO appears as a highly tunable oscillator, able to adopt period lengths from a few hours to multiple days. This is in stark contrast to the circadian timer which cannot adopt periods that are more than a few hours off from 24 hr when its limits of entrainment are tested experimentally (***Aschoff and Pohl, 1978***).

Given that DA is known to stimulate locomotor activity (***Zhou and Palmiter, 1995***), our observation of a cyclical rise in extracellular striatal DA, which is in synchrony with ultradian activity rhythms, argues that DA acts as an output of the DUO. Our finding that manipulations affecting extracellular DA levels alter oscillator period and that extracellular DA tone shows a remarkably high correlation with activity

cycle length, strongly suggests that DA is a period determinant and therefore must be an integral component of the oscillatory mechanism itself. As all period-altering manipulations directly impinge upon DA neuronal physiology, this suggests that either (i) DA neurons are the site of ultradian rhythm generation or (ii) they are a key cog in the oscillatory mechanism. The chemogenetic activation experiments indicate that the relevant DA neuronal population is located in the midbrain as selective activation of DA neurons in this region had a period lengthening effect on the ultradian activity. Future experiments will be aimed at delineating the precise DA population(s) required for the ultradian rhythm generation process and how rhythmic synchrony between neurons is maintained, if the DUO is indeed composed of a population of cellular oscillators.

Our data indicate that the circadian and ultradian locomotor rhythms are normally harmonized (for instance, ***Figure 9A***), suggesting that the circadian pacemaker and the DUO interact. Of note, extracellular DA was reported to fluctuate diurnally in the rodent striatum (***Hood et al., 2010***), which was abrogated in *Slc6a3*$^{-/-}$ mice (***Gallardo et al., 2014***). In these studies, only group averaged DA profiles were presented and DA was solely measured in circadian intact mice and rats, thus it is conceivable that any ultradian component (in WT) or infradian component (in *Slc6a3*$^{-/-}$) escaped detection. Critically, the observation that the DA levels, on average, followed a diurnal pattern with a night-time peak, together with the observation that ultradian locomotor period in female hamsters is longer in the dark vs the daily light phase (***Prendergast et al., 2012***), is consistent with the LD cycle and/or the circadian pacemaker affecting DUO rhythmicity by altering (extracellular) DA. Note however, that dopamine content in whole brain extracts from circadian-intact rats kept under LD showed a strong ultradian variation with no obvious evidence for a diurnal rhythmic component (***Scheving et al., 1968***). Also, microdialysis did not reveal a day:night difference in extracellular DA levels in the striatal NAcc region when measured in DD (***Chung et al., 2014***). This same study reported elevated DA levels and hyperactivity in mice lacking clock gene *Rev-erbα*, suggesting that the circadian clock, possibly intrinsic to DA neurons themselves, has a role in DA regulation. Indeed, knockdown of the core circadian clock component *Clock* in DA neurons of the murine VTA, increases electrical firing rate in VTA neurons and enhances locomotor response to novel objects (***Mukherjee et al., 2010***). Interestingly, we did not observe any systematic differences in ultradian locomotor period between SCNx and *Bmal1*$^{-/-}$ mice in constant darkness conditions, regardless of DAT status (***Figures 1, 2***), suggesting that extra-SCN circadian clocks have no role in DUO-mediated ultradian locomotor rhythm generation.

Our data also provide evidence that the postulated MASCO (***Tataroglu et al., 2006***) represents a long-period manifestation of the DUO as a result of methamphetamine's action on the dopamine transporter, blocking dopamine reuptake thereby increasing extracellular DA levels. Interestingly, the dopamine system has been also implicated with another behavioral timing system: the food-entrainable oscillator (FEO). This circadian independent oscillator (***Pitts et al., 2003***; ***Pendergast et al., 2009***; ***Storch and Weitz, 2009***) drives food-anticipatory locomotor activity (FAA) that emerges when food access is restricted to a few hours each day (***Mistlberger, 2011***). D1 (DRD1) and D2 (DRD2) receptor antagonists attenuate FAA additively (***Liu et al., 2012***) while pharmacological DRD2, but not DRD1, activation altered the phase of FAA (***Smit et al., 2013***). Most recently, using knockout mice, it was revealed that DRD1 but not DRD2 is necessary for the appropriate expression of FAA and that rescuing dopamine signaling selectively within the dorsal striatum was sufficient to restore FAA in dopamine-deficient mice (***Gallardo et al., 2014***). Given the links to the dopaminergic system, it will be of interest to investigate whether the DUO has a role in the temporal regulation of FAA.

While these findings argue that food cues engage the dopamine system to alter daily locomotor activity patterns, it is clear that dopamine signaling also affects food intake as mice lacking dopamine are lethargic and do not actively consume food (***Zhou and Palmiter, 1995***). Considering that the DUO is a universal driver of ultradian behavioral rhythms in mammals, the finding that ultradian bouts of running-wheel and feeding activity are co-expressed in the common vole (***van der Veen et al., 2006***) suggests that dopamine can synchronously drive food seeking and general activity, which is in line with the view that dopamine acts as a general promoter of motivated arousal.

Slc6a3$^{-/-}$ mice have been proposed as a model for schizophrenia (***Gainetdinov et al., 2001***) and the DA hypothesis of schizophrenia states that DA elevation is causal to the behavioral symptoms of this psychiatric condition. Intriguingly, circabidian (48 hr) or free-running rhythms in locomotor activity reminiscent of the behavioral patterns we detected in *Slc6a3*$^{-/-}$ or Meth-treated mice (***Figures 4F, 7C,F, 9B***) have been observed in schizophrenic subjects (***Wirz-Justice et al., 2001***; ***Wulff et al., 2012***),

suggesting that DUO dysregulation underlies the rest-activity aberrations associated with schizophrenia. Furthermore, actigraphy recordings in schizophrenic patients also revealed that Hal treatment reduces circadian/diurnal locomotor amplitude and leads to the emergence of ultradian activity bouts (**Wirz-Justice et al., 1997**, **2009**), effects we likewise observed in *Slc6a3*$^{-/-}$ mice in response to Hal. Transient DUO period lengthening might equally account for the rest-activity pattern abnormalities observed during manic episodes in bipolar disorder, which has been also associated with altered DA tone (**Berk et al., 2007**). Bipolar subjects have been reported to show rapid, 48-hr cycling between mania and depression (**Gann et al., 1993**; **Wilk and Hegerl, 2010**), one to multiple 48-hr sleep–wake cycles when switching from depression to mania (**Wehr et al., 1982**), or a long-period 'free-running' rhythm of wakefulness (**Wehr et al., 1998**). Thus schizophrenic and bipolar subjects both appear to exhibit rest-activity cycle aberrations strikingly similar to those observed in *Slc6a3*$^{-/-}$ or Meth-treated mice, suggesting that DUO dysregulation is a common indicator for these psychopathologies and perhaps even a common disease cause.

A switch to sleep–wake cycles with a period much longer than 24 hr have also been observed in subjects that were studied in temporal isolation (**Aschoff, 1965**). Because other physiological parameters such as urine secretion and core body temperature showed phase-aligned, circadian fluctuations with a period much closer to 24 hr, the subjects were considered internally desynchronized. Notably, affected subjects tend to exhibit high scores of neuroticism (**Wever, 1979**). It is therefore conceivable that the observed internal desynchronization is also due to a dysregulation of the DUO.

Materials and methods

Animals

Bmal1$^{-/-}$ (**Bunger et al., 2000**) and *DAT-Cre* (**Turiault et al., 2007**) mice were on a C57BL/6J genetic background, while *Slc6a3*$^{-/-}$ mice (**Giros et al., 1996**) were maintained on a mixed C57BL/6JxDBA/2J background (**Morice et al., 2004**). Animals were housed under LD 12 hr:12 hr unless otherwise stated. *Slc6a3*$^{-/-}$ mice were found to exhibit high attrition rates (~80%) in experiments that involve long-term locomotor activity monitoring. To increase survival, *Slc6a3*$^{-/-}$ mice were first group-housed for 1 week after transfer into the light-tight, ventilated cabinets used for activity monitoring. This was followed by 1 week of individual housing in running wheel cages with the wheels locked before commencing baseline activity recordings. During this adaptation phase and throughout the subsequent experimental period, both *Slc6a3*$^{-/-}$ mice and their wild-type littermates were provided with ad libitum chocolate flavored chow (Supreme Mini-Treats, BioServ, Flemington, NJ) as well as cotton nestlets and ample shredded corrugated card stock. This regimen significantly reduced attrition rates to less than 20% on average. 'Material and methods' were performed in accordance with the Canadian Council on Animal Care guidelines and approved by the McGill University Animal Care Committee.

Locomotor activity and core body temperature monitoring

Running wheels: Animals were individually housed in light-controlled cabinets and activity was recorded continuously (ClockLab, Actimetrics, Wilmette, IL). Actograms, displaying binned running wheel revolutions per 6 min (0.1 hr), and the associated Lomb-Scargle periodograms, displaying amplitude, were generated using ClockLab software. *Telemetry:* Animals were individually housed in standard cages placed atop energizer/receiver units (ER-4000, Starr Life Science Corp., Oakmont, PA). 1 week prior to data collection, electromagnetic induction powered telemetry probes (G2 E-mitter, Starr Life Science Corp.) were implanted intraperitoneally. Locomotion, measured in counts per minute, and core body temperature (BT, in °C), was collected in 6-min bins (0.1 hr) using Vitalview software (Starr Life Science Corp.). BT data was exported into Clocklab to generate actogram-style data displays with tick mark height corresponding to temperatures from 34–38°C; for waveform generation, averaging, and 2 hr recursive smoothing, locomotor activity and BT data was exported to Excel (Microsoft, Redmond, WA). Matlab (Mathworks, Natick, MA) was used for ribbon plot generation, and low pass Butterworth filtering (1 hr).

Pharmacology

Stock solutions of (+)-Methamphetamine hydrochloride (100 mg/l, Sigma-Aldrich, St. Louis, MO), D-Amphetamine hemisulfate (100 mg/l, Sigma-Alrdich), Haloperidol (11.25 mg/ml, Sigma-Aldrich) and Clozapine-N-Oxide (15 mg/l, National Institute of Mental Health, Bethesda, MD) were prepared

using distilled water. Methamphetamine solutions were adjusted to pH 7 using sodium hydroxide and haloperidol was dissolved by stirring at 40°C.

Chronic methamphetamine infusion

Mice were continuously infused with methamphetamine (0.6 mg/day in 0.9% saline) for 14 days using subcutaneously implanted osmotic minipumps (Alzet Model 1002, Durect Corp., Cupertino CA). Pumps were fitted with a 65-mm polyvinyl chloride catheter (0.69-mm inner diameter, Plastics One, Roanoke, VA) and backfilled with saline to allow for an approximately 4 day post-surgery recovery prior to drug exposure.

In situ hybridization

In situ hybridization was performed as previously described (***Kraves and Weitz, 2006***). Briefly, following decapitation, brains were removed and quickly frozen at −80°C. Serial coronal hypothalamic brain slices (25 μm) were collected using a cryostat and stored at −80°C until hybridization. Sections were hybridized overnight at 60°C to a digoxigenin-labeled riboprobe targeting the coding regions of mouse *Per1* (nucleotides 579–1478 of the *Per1* mRNA, Genbank, NM_011065.4).

SCN lesions

Electrolytic lesions of the suprachiasmatic nucleus were performed as described (***Storch et al., 2007***). Briefly, an electrode (RNE-300X, Rhodes Medical Instruments, Woodland Hills, CA) was lowered through a hole drilled in the skull at the mid-sagittal sinus according to stereotaxic coordinates (AP −0.25 mm, DV −6.00 mm from Bregma) and a constant current (2 mA, 10 s; D.C. Constant Lesion Maker, Grass Instruments, Quincy, MA) was applied. Only mice with behavioral circadian arrhythmia (in the 20–28-hr range assessed by Lomb-Scargle periodogram analysis) and subsequent post-mortem histological verification using DAPI staining mounting medium (Vectashield, Vector Labs) were included for analysis. This represented approximately 50% of lesioned mice when averaged across all studies.

In vivo microdialysis

1 week prior to sampling, mice were stereotaxically implanted with a guide cannula (C312G/spc 2.5 mm below pedestal, Plastics One), targeting the striatum (AP +1.10 mm, DV +5.5 mm, ML +0.9 mm). 1 day prior to dialysis, mice were transferred to a beam break monitor (VersaMax, Omnitech Electronics, Columbus, OH) and tethered using a dummy probe assembly under constant, dim, red light (<5lux). On the day of sampling, in-house recording probes (***Lupinsky et al., 2010***), were connected to a central swivel (Model 72-0000, Harvard Apparatus, Holliston, MA). Artificial cerebrospinal fluid was delivered via syringe pump (Model 403, CMA Microdialysis, Kista, Sweden) at a flow rate of 1.0 μl/min. This procedure was similarly followed to test the effects of CNO in AAV-transduced *DAT-Cre* mice where intraperitoneal injections of 1 mg/kg CNO were given after approximately 1 hr of baseline sampling. Samples were collected every 20 min for 4 hr and immediately admixed with 1 μl of perchloric acid. DA and DOPAC content in the dialysate were determined by high-performance liquid chromatography with electrochemical detection (HPLC-EC) as described (***Domenger et al., 2012***). Chromatographic peak analysis was conducted using CoulArray software (ESA Inc., Chelmsford, MA). After data collection, brains were removed, sectioned, and stained with hematoxylin to verify probe placement. Locomotor activity data, recorded as the number of beam breaks per minute, were exported into Matlab (Mathworks) for waveform generation, 20-min binning, cosinor modeling, and false discovery rate analysis.

DA/DOPAC content determination in tissue extracts

After decapitation, brains were quickly removed and frozen (−80°C). 320 μm coronal brain slices were obtained by cryosectioning and then microdissected (1- or 2-mm diameter sample corers, Fine Science Tools Inc., Foster City, CA) to obtain tissue of the CPu (2 mm), NAcc (1 mm), and SN/VTA midbrain (2 mm) regions (see ***Figure 6G*** for punch location). DA/DOPAC was quantified as previously described (***Domenger et al., 2012***). Briefly, individual punches from each region were homogenized in 45 μl perchloric acid (0.25 M) to which 15 μl of a 100 ng/ml solution of 3,4-dihydroxybenzylamine was added, which served as the internal standard. Concentrations were determined by HPLC-EC. After perchloric acid extraction, the protein containing pellets were reconstituted in 0.1 N sodium hydroxide for protein quantification (Pierce BCA Kit, ThermoFisher Scientific, Waltham, MA).

Chemogenetics

Mice were anaesthetized with isofluorane and placed in a stereotaxic aparatus (David Kopf Instruments). Recombinant *AAV8-DIO-h3MDq-mCherry* (**Krashes et al., 2011**) (titer = 3×10^{12} genomes copies per ml, UNC Vector Core Services, Chapel Hill, NC) was bilaterally injected into to the VTA/SN area (coordinates from bregma: AP: −3.44 mm, DV: −4.40 mm, L: ±0.48 mm) (**Tsai et al., 2009**) through a cannula (33 gauge, Plastics One) at a flow rate of 0.05 μl/min for 10 min (0.5 μl total volume per side) using a syringe pump (Harvard Apparatus). Mice were subsequently maintained in individual housing for at least 2 weeks prior to CNO treatment.

Immunohistochemistry

Immunostaining was performed as previously described (**Chu et al., 2013**) using cryoprotected (30% Sucrose), free-floating coronal sections from fixed brains collected after intra-cardiac perfusion (4% paraformaldehyde) and cut at 40 microns using a cryostat (Leica, Solms, Germany). For fluorescent labeling, sections were incubated overnight with primary antibodies for mCherry (rabbit anti-RFP, 1:1000, Rockland, Limerick, PA) to enhance detection of the mCherry expression and antibodies for tyrosine hydroxylase (mouse anti-TH, 1:1000, EMD Millipore, Etibicoke, Canada). This was followed by 2 hr incubation with secondary antibodies (anti-rabbit Alexa Fluor 567 and anti-mouse Alexa Fluor 647, 1:250, Life Technologies, Carlsbad, CA) after which sections were mounted on superfrost slides (VWR, Radnor, PA), coverslipped with Vectashield mounting medium (Vector Labs, Burlington, Canada) and imaged by fluorescence microscopy (AxioObserver Z1, Zeiss, Jena, Germany). Quantification of co-expression was performed on a single, medial, midbrain section from each behaviorally responsive mouse using the cell counter plugin of ImageJ (National Institute of Health, Bethesda, MD) in order to determine the percentage of all TH-positive cells that also expressed the hM3Dq--mCherry fusion protein (N = 6 per genotype).

Statistical analysis

One-way and repeated measures ANOVA, t-tests, and linear regression analyses were performed using Prism 5 (GraphPad, La Jolla, CA). Planned comparisons for the repeated measures were carried out to determine significant period changes between subsequent measurements. Bonferroni correction was used for post-hoc testing of individual group differences.

False discovery rate (FDR) and cosinor determination

To evaluate the probabilistic significance of the least-square fit between the DA trace and the cosinor model that was computed based on the animal's locomotor activity oscillations, a false discovery rate approach was employed (**Storch et al., 2007**). For cosinor computation, we first determined the locomotor period of *Bmal1*$^{-/-}$ mice by Lomb-Scargle periodogram analysis of 20 hr of locomotor recordings following dialysate sampling. The determined period was then used as an input parameter for a least-squares cosinor analysis (**Nelson et al., 1979**) using a custom script and function written for Matlab (https://github.com/storchlab/Cosinor-FDR.git). This procedure generated a best-fit (co)sine wave modeling the activity time-series. This model wave was then used to assess the concordance between the extracellular DA fluctuations and locomotor oscillations. To calculate FDR, we randomly permutated the temporal order of the measured DA concentrations 100,000 times, and assessed its fit to the cosinor. The permutation procedure will degrade the fit of any truly rhythmic signal but will have little to no effect on noisy or flat profiles. The probability of false discovery was calculated as the percentage of randomly permuted traces that show an equal or better fit to the cosinor than the measured DA fluctuation.

Butterworth filtering

Low pass filtering (1-hr cut-off) of raw data was conducted using a Butterworth zero-phase filter (**Butterworth, 1930**) in Matlab. This allowed for better visualization of ultradian frequencies in the >1-hr range, in a manner similar to recursive smoothing but without phase distortion.

Period determination

Ultradian rhythms show variability in amplitude and period. We thus used Lomb-Scargle periodogram analysis (Clocklab) to estimate ultradian period length as this method is relatively tolerant to noisy data and data gaps, which can result from intermittent ultradian rhythm expression (**Press and Rybicki, 1989**). Unless otherwise stated, period length was determined by identifying the highest peak above the significance threshold ($\alpha = 0.01$) in the Lomb-Scargle periodogram. Subsequent

plotting of an actogram at the determined modulus was conducted in each case in order to visually confirm rhythmicity and rule out the false identification of harmonic frequencies or side-lobes due to leakage (***Van Dongen et al., 1999***).

Continuous wavelet transforms

To visualize and quantify the dynamics of period and amplitude in circadian incompetent mice, continuous wavelet transforms using the Morlet wavelet (***Goupillaud et al., 1984***) were performed for the 1–12-hr range using custom algorithms written in Matlab (https://github.com/storchlab/CWT.git). Locomotor activity data of seven consecutive days per individual SCN-lesioned or *Bmal1*$^{-/-}$ mouse was used. For each spectrogram, a ridge identifying the peak amplitude of the spectrum was generated as described (***Leise, 2013***) and the associated values were used to calculate the average and standard deviation of the dominant ultradian frequency for each animal.

Amplitude spectral density (ASD)

To determine the prevalence of oscillations in a given period range we calculated the area under the curve of all significantly rhythmic periodicities ($\alpha = 0.01$) in the Lomb-Scargle periodogram, that is, the spectral density. For normalization, the obtained spectral density was divided by the total significant spectral density in the 0–40-hr range and expressed as percentage.

Acknowledgements

We thank A Villemain, E Vigneault, and M-E Desaulnier for mouse-colony maintenance and F Tronche for the *DAT-Cre* mice. This work was supported by a NSERC discovery grant (K-FS), CIHR operating grants (K-FS and AG), a CIHR New Investigator Award (K-FS), by the Graham Boeckh Foundation (BG) and a Canada Research Chair (BG).

Additional information

Funding

Funder	Grant reference number	Author
Canadian Institutes of Health Research	Operating Grants	Alain Gratton, Kai-Florian Storch
Natural Sciences and Engineering Research Council of Canada	Discovery Grant	Kai-Florian Storch
Graham Boeckh Foundation	Graham Boeckh Chair in Schizophrenia	Bruno Giros
Canadian Institutes of Health Research	New Investigator Award	Kai-Florian Storch
Canadian Institutes of Health Research	Canada Research Chair	Bruno Giros

The funders had no role in study design, data collection and interpretation, or the decision to submit the work for publication.

Author contributions

IDB, K-FS, Conception and design, Acquisition of data, Analysis and interpretation of data, Drafting or revising the article; LZ, LM, Acquisition of data, Analysis and interpretation of data; MVK, AG, BG, Conception and design, Drafting or revising the article, Contributed unpublished essential data or reagents

Author ORCIDs

Ian D Blum, http://orcid.org/0000-0001-6999-8364

Ethics

Animal experimentation: All experimental procedures were performed in accordance with the Canadian Council on Animal Care guidelines and approved by the McGill University Animal Care Committee (animal use protocol #2010-5945).

References

Alexander GM, Rogan SC, Abbas AI, Armbruster BN, Pei Y, Allen JA, Nonneman RJ, Hartmann J, Moy SS, Nicolelis MA, McNamara JO, Roth BL. 2009. Remote control of neuronal activity in transgenic mice expressing evolved G protein-coupled receptors. *Neuron* **63**:27–39. doi: 10.1016/j.neuron.2009.06.014.

Aschoff J. 1965. Circadian rhythms in man. *Science* **148**:1427–1432. doi: 10.1126/science.148.3676.1427.

Aschoff J, Pohl H. 1978. Phase relations between a circadian rhythm and its zeitgeber within the range of entrainment. *Die Naturwissenschaften* **65**:80–84. doi: 10.1007/BF00440545.

Bengel D, Murphy DL, Andrews AM, Wichems CH, Feltner D, Heils A, Mössner R, Westphal H, Lesch KP. 1998. Altered brain serotonin homeostasis and locomotor insensitivity to 3, 4-methylenedioxymethamphetamine ('Ecstasy') in serotonin transporter-deficient mice. *Molecular Pharmacology* **53**:649–655.

Berk M, Dodd S, Kauer-Sant'anna M, Malhi GS, Bourin M, Kapczinski F, Norman T. 2007. Dopamine dysregulation syndrome: implications for a dopamine hypothesis of bipolar disorder. *Acta Psychiatrica Scandinavica Supplementum*:41–49.

Brown RE, Basheer R, McKenna JT, Strecker RE, McCarley RW. 2012. Control of sleep and wakefulness. *Physiological Reviews* **92**:1087–1187. doi: 10.1152/physrev.00032.2011.

Bunger MK, Wilsbacher LD, Moran SM, Clendenin C, Radcliffe LA, Hogenesch JB, Simon MC, Takahashi JS, Bradfield CA. 2000. Mop3 is an essential component of the master circadian pacemaker in mammals. *Cell* **103**:1009–1017. doi: 10.1016/S0092-8674(00)00205-1.

Butterworth S. 1930. On the theory of filter amplifiers. *Experimental Wireless* **7**:536–541.

Chu A, Zhu L, Blum ID, Mai O, Leliavski A, Fahrenkrug J, Oster H, Boehm U, Storch KF. 2013. Global but not gonadotrope-specific disruption of Bmal1 abolishes the luteinizing hormone surge without affecting ovulation. *Endocrinology* **154**:2924–2935. doi: 10.1210/en.2013-1080.

Chung S, Lee EJ, Yun S, Choe D, Park SB, Son HJ, Kim KS, Dluzen DE, Lee I, Hwang O, Son GH, Kim K. 2014. Impact of circadian nuclear receptor REV-ERBα on midbrain dopamine production and mood regulation. *Cell* **157**:858–868. doi: 10.1016/j.cell.2014.03.039.

Daan S, Slopsema S. 1978. Short-term rhythms in foraging behaviour of the common vole, Microtus arvalis. *Journal of Comparative Physiology* **127**:215–227. doi: 10.1007/BF01350112.

de la Iglesia HO, Cambras T, Schwartz WJ, Diez-Noguera A. 2004. Forced desynchronization of dual circadian oscillators within the rat suprachiasmatic nucleus. *Current Biology* **14**:796–800. doi: 10.1016/j.cub.2004.04.034.

Domenger D, Dea D, Theroux L, Moquin L, Gratton A, Poirier J. 2012. The MPTP neurotoxic lesion model of Parkinson's disease activates the apolipoprotein E cascade in the mouse brain. *Experimental Neurology* **233**:513–522. doi: 10.1016/j.expneurol.2011.11.031.

Dowse HB, Hall JC, Ringo JM. 1987. Circadian and ultradian rhythms in period mutants of *Drosophila melanogaster*. *Behavior Genetics* **17**:19–35. doi: 10.1007/BF01066008.

Dowse H, Umemori J, Koide T. 2010. Ultradian components in the locomotor activity rhythms of the genetically normal mouse, Mus musculus. *The Journal of Experimental Biology* **213**:1788–1795. doi: 10.1242/jeb.038877.

Gainetdinov RR, Jones SR, Fumagalli F, Wightman RM, Caron MG. 1998. Re-evaluation of the role of the dopamine transporter in dopamine system homeostasis. *Brain Research Brain Research Reviews* **26**:148–153. doi: 10.1016/S0165-0173(97)00063-5.

Gainetdinov RR, Mohn AR, Caron MG. 2001. Genetic animal models: focus on schizophrenia. *Trends in Neurosciences* **24**:527–533. doi: 10.1016/S0166-2236(00)01886-5.

Gallardo CM, Darvas M, Oviatt M, Chang CH, Michalik M, Huddy TF, Meyer EE, Shuster SA, Aguayo A, Hill EM, Kiani K, Ikpeazu J, Martinez JS, Purpura M, Smit AN, Patton DF, Mistlberger RE, Palmiter RD, Steele AD. 2014. Dopamine receptor 1 neurons in the dorsal striatum regulate food anticipatory circadian activity rhythms in mice. *eLife* **3**:e03781. doi: 10.7554/eLife.03781.

Gann H, Riemann D, Hohagen F, Strauss LG, Dressing H, Müller WE, Berger M. 1993. 48-hour rapid cycling: results of psychopathometric, polysomnographic, PET imaging and neuro-endocrine longitudinal investigations in a single case. *Journal of Affective Disorders* **28**:133–140. doi: 10.1016/0165-0327(93)90042-I.

Gerkema MP, Groos GA, Daan S. 1990. Differential elimination of circadian and ultradian rhythmicity by hypothalamic lesions in the common vole, Microtus arvalis. *Journal of Biological Rhythms* **5**:81–95. doi: 10.1177/074873049000500201.

Gerkema MP, van der Leest F. 1991. Ongoing ultradian activity rhythms in the common vole, Microtus arvalis, during deprivations of food, water and rest. *Journal of Comparative Physiology A, Sensory, Neural, and Behavioral Physiology* **168**:591–597. doi: 10.1007/BF00215081.

Gerkema MP, Verhulst S. 1990. Warning against an unseen predator: a functional aspect of synchronous feeding in the common vole, Microtus arvalis. *Animal Behaviour* **40**:1169–1178. doi: 10.1016/S0003-3472(05)80183-6.

Giros B, Jaber M, Jones SR, Wightman RM, Caron MG. 1996. Hyperlocomotion and indifference to cocaine and amphetamine in mice lacking the dopamine transporter. *Nature* **379**:606–612. doi: 10.1038/379606a0.

Goodwin JS, Larson GA, Swant J, Sen N, Javitch JA, Zahniser NR, De Felice LJ, Khoshbouei H. 2009. Amphetamine and methamphetamine differentially affect dopamine transporters in vitro and in vivo. *The Journal of Biological Chemistry* **284**:2978–2989. doi: 10.1074/jbc.M805298200.

Goupillaud P, Grossmann A, Morlet J. 1984. Cycle-octave and related transforms in seismic signal analysis. *Geoexploration* **23**:85–102. doi: 10.1016/0016-7142(84)90025-5.

Honma KI, Hiroshige T. 1978. Endogenous ultradian rhythms in rats exposed to prolonged continuous light. *The American Journal of Physiology* **235**:R250–R256.
Honma K, Honma S, Hiroshige T. 1986. Disorganization of the rat activity rhythm by chronic treatment with methamphetamine. *Physiology & Behavior* **38**:687–695. doi: 10.1016/0031-9384(86)90265-9.
Honma K, Honma S, Hiroshige T. 1987. Activity rhythms in the circadian domain appear in suprachiasmatic nuclei lesioned rats given methamphetamine. *Physiology & Behavior* **40**:767–774. doi: 10.1016/0031-9384(87)90281-2.
Hood S, Cassidy P, Cossette MP, Weigl Y, Verwey M, Robinson B, Stewart J, Amir S. 2010. Endogenous dopamine regulates the rhythm of expression of the clock protein PER2 in the rat dorsal striatum via daily activation of D2 dopamine receptors. *The Journal of Neuroscience* **30**:14046–14058. doi: 10.1523/JNEUROSCI.2128-10.2010.
Howell LL, Kimmel HL. 2008. Monoamine transporters and psychostimulant addiction. *Biochemical Pharmacology* **75**:196–217. doi: 10.1016/j.bcp.2007.08.003.
Ibuka N, Inouye SI, Kawamura H. 1977. Analysis of sleep-wakefulness rhythms in male rats after suprachiasmatic nucleus lesions and ocular enucleation. *Brain Research* **122**:33–47. doi: 10.1016/0006-8993(77)90660-6.
Ichikawa J, Meltzer HY. 1991. Differential effects of repeated treatment with haloperidol and clozapine on dopamine release and metabolism in the striatum and the nucleus accumbens. *The Journal of Pharmacology and Experimental Therapeutics* **256**:348–357.
Krashes MJ, Koda S, Ye C, Rogan SC, Adams AC, Cusher DS, Maratos-Flier E, Roth BL, Lowell BB. 2011. Rapid, reversible activation of AgRP neurons drives feeding behavior in mice. *The Journal of Clinical Investigation* **121**:1424–1428. doi: 10.1172/JCI46229.
Kraves S, Weitz CJ. 2006. A role for cardiotrophin-like cytokine in the circadian control of mammalian locomotor activity. *Nature Neuroscience* **9**:212–219. doi: 10.1038/nn1633.
Lane RF, Blaha CD. 1987. Chronic haloperidol decreases dopamine release in striatum and nucleus accumbens in vivo: depolarization block as a possible mechanism of action. *Brain Research Bulletin* **18**:135–138. doi: 10.1016/0361-9230(87)90042-6.
Leise TL. 2013. Wavelet analysis of circadian and ultradian behavioral rhythms. *Journal of Circadian Rhythms* **11**:5. doi: 10.1186/1740-3391-11-5.
Liu YY, Liu TY, Qu WM, Hong ZY, Urade Y, Huang ZL. 2012. Dopamine is involved in food-anticipatory activity in mice. *Journal of Biological Rhythms* **27**:398–409. doi: 10.1177/0748730412455913.
Lupinsky D, Moquin L, Gratton A. 2010. Interhemispheric regulation of the medial prefrontal cortical glutamate stress response in rats. *The Journal of Neuroscience* **30**:7624–7633. doi: 10.1523/JNEUROSCI.1187-10.2010.
Mistlberger RE. 2011. Neurobiology of food anticipatory circadian rhythms. *Physiology & Behavior* **104**:535–545. doi: 10.1016/j.physbeh.2011.04.015.
Mohawk JA, Baer ML, Menaker M. 2009. The methamphetamine-sensitive circadian oscillator does not employ canonical clock genes. *Proceedings of the National Academy of Sciences of USA* **106**:3519–3524. doi: 10.1073/pnas.0813366106.
Morice E, Denis C, Giros B, Nosten-Bertrand M. 2004. Phenotypic expression of the targeted null-mutation in the dopamine transporter gene varies as a function of the genetic background. *The European Journal of Neuroscience* **20**:120–126. doi: 10.1111/j.1460-9568.2004.03465.x.
Mukherjee S, Coque L, Cao JL, Kumar J, Chakravarty S, Asaithamby A, Graham A, Gordon E, Enwright JF III, DiLeone RJ, Birnbaum SG, Cooper DC, McClung CA. 2010. Knockdown of Clock in the ventral tegmental area through RNA interference results in a mixed state of mania and depression-like behavior. *Biological Psychiatry* **68**:503–511. doi: 10.1016/j.biopsych.2010.04.031.
Nelson W, Tong YL, Lee JK, Halberg F. 1979. Methods for cosinor-rhythmometry. *Chronobiologia* **6**:305–323.
Parmentier R, Ohtsu H, Djebbara-Hannas Z, Valatx JL, Watanabe T, Lin JS. 2002. Anatomical, physiological, and pharmacological characteristics of histidine decarboxylase knock-out mice: evidence for the role of brain histamine in behavioral and sleep-wake control. *The Journal of Neuroscience* **22**:7695–7711.
Paxinos G, Franklin KBJ. 2001. *The Mouse Brain in Stereotaxic Coordinates*. New York, NY: Academic Press.
Pendergast JS, Nakamura W, Friday RC, Hatanaka F, Takumi T, Yamazaki S. 2009. Robust food anticipatory activity in BMAL1-deficient mice. *PLOS ONE* **4**:e4860. doi: 10.1371/journal.pone.0004860.
Pitts S, Perone E, Silver R. 2003. Food-entrained circadian rhythms are sustained in arrhythmic Clk/Clk mutant mice. *American Journal of Physiology Regulatory, Integrative and Comparative Physiology* **285**:R57–R67.
Prendergast BJ, Cisse YM, Cable EJ, Zucker I. 2012. Dissociation of ultradian and circadian phenotypes in female and male *Siberian hamsters*. *Journal of Biological Rhythms* **27**:287–298. doi: 10.1177/0748730412448618.
Press WH, Rybicki GB. 1989. Fast algorithm for spectral analysis of unevenly sampled data. *The Astrophysical Journal* **338**:277–280. doi: 10.1086/167197.
Rivkees SA. 2003. Developing circadian rhythmicity in infants. *Pediatrics* **112**:373–381. doi: 10.1542/peds.112.2.373.
Ruis JF, Rietveld WJ, Buys PJ. 1987. Effects of suprachiasmatic nuclei lesions on circadian and ultradian rhythms in body-temperature in ocular enucleated rats. *Journal of Interdisciplinary Cycle Research* **18**:259–273. doi: 10.1080/09291018709359952.
Rusak B. 1977. The role of the suprachiasmatic nuclei in the generation of circadian rhythms in the golden hamster, Mesocricetus auratus. *Journal of Comparative Physiology* **118**:145–164. doi: 10.1007/BF00611819.
Scheving LE, Harrison WH, Gordon P, Pauly JE. 1968. Daily fluctuation (circadian and ultradian) in biogenic amines of the rat brain. *The American Journal of Physiology* **214**:166–173.
Schibler U. 2008. The mammalian circadian timekeeping system. In: Lloyd D, Rossi E, editors. *Ultradian Rhythms From Molecules to Mind*. Netherlands: Springer. p. 261–279.

Schmitz Y, Benoit-Marand M, Gonon F, Sulzer D. 2003. Presynaptic regulation of dopaminergic neurotransmission. *Journal of Neurochemistry* **87**:273–289. doi: 10.1046/j.1471-4159.2003.02050.x.
Seki Y, Tanimura T. 2014. Ultradian rhythm unmasked in the Pdf clock mutant of *Drosophila*. *Journal of Biosciences* **39**:585–594. doi: 10.1007/s12038-014-9450-z.
Smit AN, Patton DF, Michalik M, Opiol H, Mistlberger RE. 2013. Dopaminergic regulation of circadian food anticipatory activity rhythms in the rat. *PLOS ONE* **8**:e82381. doi: 10.1371/journal.pone.0082381.
Storch KF, Weitz CJ. 2009. Daily rhythms of food-anticipatory behavioral activity do not require the known circadian clock. *Proceedings of the National Academy of Sciences of USA* **106**:6808–6813. doi: 10.1073/pnas.0902063106.
Storch KF, Paz C, Signorovitch J, Raviola E, Pawlyk B, Li T, Weitz CJ. 2007. Intrinsic circadian clock of the mammalian retina: importance for retinal processing of visual information. *Cell* **130**:730–741. doi: 10.1016/j.cell.2007.06.045.
Strange PG. 2008. Antipsychotic drug action: antagonism, inverse agonism or partial agonism. *Trends in Pharmacological Sciences* **29**:314–321. doi: 10.1016/j.tips.2008.03.009.
Sulzer D. 2011. How addictive drugs disrupt presynaptic dopamine neurotransmission. *Neuron* **69**:628–649. doi: 10.1016/j.neuron.2011.02.010.
Tannenbaum GS, Martin JB. 1976. Evidence for an endogenous ultradian rhythm governing growth hormone secretion in the rat. *Endocrinology* **98**:562–570. doi: 10.1210/endo-98-3-562.
Tataroglu O, Davidson AJ, Benvenuto LJ, Menaker M. 2006. The methamphetamine-sensitive circadian oscillator (MASCO) in mice. *Journal of Biological Rhythms* **21**:185–194. doi: 10.1177/0748730406287529.
Thomas SA, Palmiter RD. 1997. Disruption of the dopamine beta-hydroxylase gene in mice suggests roles for norepinephrine in motor function, learning, and memory. *Behavioral Neuroscience* **111**:579–589. doi: 10.1037/0735-7044.111.3.579.
Tsai HC, Zhang F, Adamantidis A, Stuber GD, Bonci A, de Lecea L, Deisseroth K. 2009. Phasic firing in dopaminergic neurons is sufficient for behavioral conditioning. *Science* **324**:1080–1084. doi: 10.1126/science.1168878.
Turiault M, Parnaudeau S, Milet A, Parlato R, Rouzeau JD, Lazar M, Tronche F. 2007. Analysis of dopamine transporter gene expression pattern – generation of DAT-iCre transgenic mice. *The FEBS Journal* **274**:3568–3577. doi: 10.1111/j.1742-4658.2007.05886.x.
van der Veen DR, Minh NL, Gos P, Arneric M, Gerkema MP, Schibler U. 2006. Impact of behavior on central and peripheral circadian clocks in the common vole Microtus arvalis, a mammal with ultradian rhythms. *Proceedings of the National Academy of Sciences of USA* **103**:3393–3398. doi: 10.1073/pnas.0507825103.
Van Dongen HP, Olofsen E, Van Hartevelt JH, Kruyt EW. 1999. A procedure of multiple period searching in unequally spaced time-series with the Lomb–Scargle method. *Biological Rhythm Research* **30**:149–177. doi: 10.1076/brhm.30.2.149.1424.
van Oort BE, Tyler NJ, Gerkema MP, Folkow L, Stokkan KA. 2007. Where clocks are redundant: weak circadian mechanisms in reindeer living under polar photic conditions. *Die Naturwissenschaften* **94**:183–194. doi: 10.1007/s00114-006-0174-2.
Vitaterna MH, King DP, Chang AM, Kornhauser JM, Lowrey PL, McDonald JD, Dove WF, Pinto LH, Turek FW, Takahashi JS. 1994. Mutagenesis and mapping of a mouse gene, Clock, essential for circadian behavior. *Science* **264**:719–725. doi: 10.1126/science.8171325.
Wang S, Tan Y, Zhang JE, Luo M. 2013. Pharmacogenetic activation of midbrain dopaminergic neurons induces hyperactivity. *Neuroscience Bulletin* **29**:517–524. doi: 10.1007/s12264-013-1327-x.
Wehr TA, Goodwin FK, Wirz-Justice A, Breitmaier J, Craig C. 1982. 48-hour sleep-wake cycles in manic-depressive illness: naturalistic observations and sleep deprivation experiments. *Archives of General Psychiatry* **39**:559–565. doi: 10.1001/archpsyc.1982.04290050037008.
Wehr TA, Turner EH, Shimada JM, Lowe CH, Barker C, Leibenluft E. 1998. Treatment of a rapidly cycling bipolar patient by using extended bed rest and darkness to stabilize the timing and duration of sleep. *Biological Psychiatry* **43**:822–828. doi: 10.1016/S0006-3223(97)00542-8.
Welsh D, Richardson GS, Dement WC. 1988. Effect of running wheel availability on circadian patterns of sleep and wakefulness in mice. *Physiology & Behavior* **43**:771–777. doi: 10.1016/0031-9384(88)90375-7.
Wever RA. 1979. In: Schaefer KE, editor. *The Circadian System of Man: Results of Experiments Under Temporal Isolation*. New York: Springer-Verlag Berlin Heildelberg.
White FJ, Wang RY. 1983. Comparison of the effects of chronic haloperidol treatment on A9 and A10 dopamine neurons in the rat. *Life Sciences* **32**:983–993. doi: 10.1016/0024-3205(83)90929-3.
Wilk K, Hegerl U. 2010. Time of mood switches in ultra-rapid cycling disorder: a brief review. *Psychiatry Research* **180**:1–4. doi: 10.1016/j.psychres.2009.08.011.
Wirz-Justice A, Bromundt V, Cajochen C. 2009. Circadian disruption and psychiatric disorders: the importance of entrainment. *Sleep Medicine Clinics* **4**:273–284. doi: 10.1016/j.jsmc.2009.01.008.
Wirz-Justice A, Cajochen C, Nussbaum P. 1997. A schizophrenic patient with an arrhythmic circadian rest-activity cycle. *Psychiatry Research* **73**:83–90. doi: 10.1016/S0165-1781(97)00117-0.
Wirz-Justice A, Haug HJ, Cajochen C. 2001. Disturbed circadian rest-activity cycles in schizophrenia patients: an effect of drugs? *Schizophrenia Bulletin* **27**:497–502. doi: 10.1093/oxfordjournals.schbul.a006890.
Wisor JP, Nishino S, Sora I, Uhl GH, Mignot E, Edgar DM. 2001. Dopaminergic role in stimulant-induced wakefulness. *The Journal of Neuroscience* **21**:1787–1794.
Wulff K, Dijk DJ, Middleton B, Foster RG, Joyce EM. 2012. Sleep and circadian rhythm disruption in schizophrenia. *The British Journal of Psychiatry* **200**:308–316. doi: 10.1192/bjp.bp.111.096321.

Xu F, Gainetdinov RR, Wetsel WC, Jones SR, Bohn LM, Miller GW, Wang YM, Caron MG. 2000. Mice lacking the norepinephrine transporter are supersensitive to psychostimulants. *Nature Neuroscience* **3**:465–471. doi: 10.1038/74839.

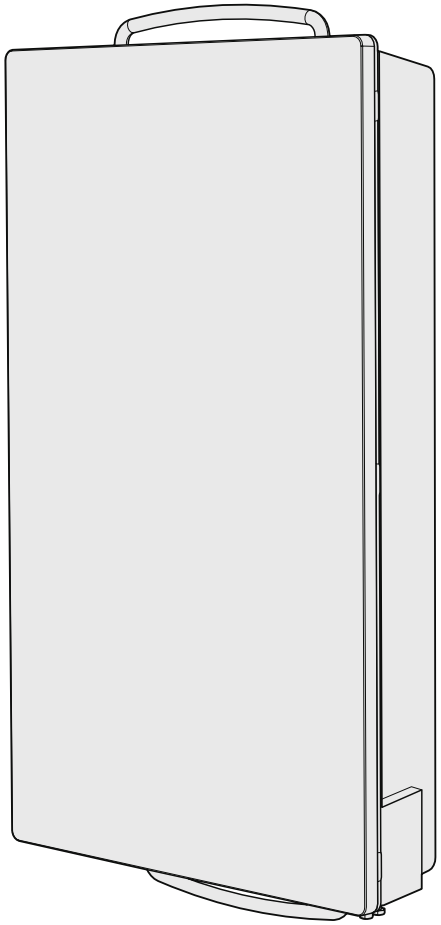
Zhao ZQ, Scott M, Chiechio S, Wang JS, Renner KJ, RW Gereau IV, Johnson RL, Deneris ES, Chen ZF. 2006. Lmx1b is required for maintenance of central serotonergic neurons and mice lacking central serotonergic system exhibit normal locomotor activity. *The Journal of Neuroscience* **26**:12781–12788. doi: 10.1523/JNEUROSCI.4143-06.2006.

Zhou QY, Palmiter RD. 1995. Dopamine-deficient mice are severely hypoactive, adipsic, and aphagic. *Cell* **83**:1197–1209. doi: 10.1016/0092-8674(95)90145-0.

Install Antenna Integrated Radio Units

AIR 6419

Installation Instructions



Ge23266B

237/1531-LZA 701 6001/1 Uen S

Copyright

© Ericsson AB 2021, 2022. All rights reserved. No part of this document may be reproduced in any form without the written permission of the copyright owner.

Disclaimer

The contents of this document are subject to revision without notice due to continued progress in methodology, design and manufacturing. Ericsson shall have no liability for any error or damage of any kind resulting from the use of this document.

Trademark List

All trademarks mentioned herein are the property of their respective owners. These are shown in the document Trademark Information.

Contents

1	**Install AIR 6419**	**1**
1.1	Safety and Operation Requirements	4
2	**Unpack and Check Materials**	**5**
2.1	Handle AIR Unit	7
3	**Pole Installation**	**10**
3.1	Prepare AIR Unit for Single Unit Pole Installation	10
3.2	Install SFP Module	13
3.3	Install RAE Unit (Optional)	15
3.4	Assemble DC Power Connector to Shielded Cable	16
3.5	Install AIR Unit on a Pole	21
4	**Wall Installation**	**26**
4.1	Prepare AIR Unit for Single Unit Wall Installation	26
4.2	Install SFP Module	29
4.3	Install RAE Unit (Optional)	32
4.4	Assemble DC Power Connector to Shielded Cable	33
4.5	Install AIR Unit on a Wall	38
5	**Connect Cables**	**42**
5.1	Hoist Cables	42
5.2	Connect Grounding Cable	43
5.3	Connect Optical Cables	44
5.4	Connect EC Light Cable (Optional)	48
5.5	Connect RAE Cable (Optional)	48
5.6	Connect External Alarm Cable (Optional)	49
5.7	Connect TX Monitor Cable (Optional)	50
5.8	Connect Power Cable	51
6	**Verify Installation**	**53**
7	**Power Up AIR**	**55**
8	**Perform Concluding Routines**	**57**
8.1	Environment	57

1 Install AIR 6419

> A work process for the installation of AIR 6419 and details about preparations to be made before starting the installation of the unit are given.

An animated installation procedure is also available at https://play.ericsson.net/media/t/1_x2g22ojt.

Work Process

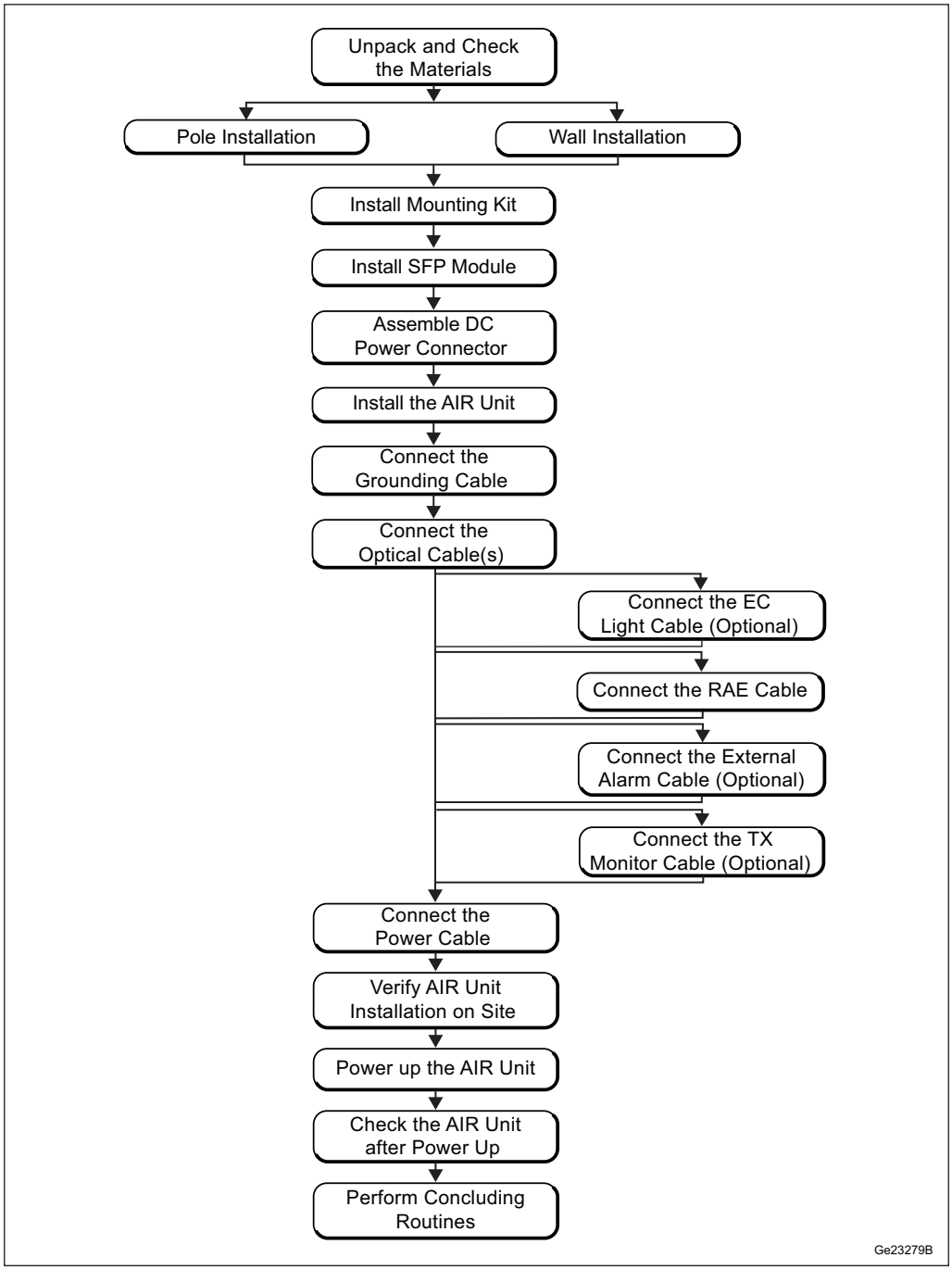


Figure 1 Radio Installation Work Process

Before Going to Site

Before going to site, ensure the following:

— Site access permission has been received.

— Following documents are read and understood:

- Personal Health and Safety Information
- System Safety Information
- Transportation and Storage

— All required tools are available.

The tools listed in this document are only examples. Tools in different areas and regions can look different than those listed here.

Depending on the installation and environment, different tools are required. It is important to do the following:

- Use tethered tools when working at heights.

Before Starting Installation

Before starting the installation, ensure the following:

— The site is prepared according to the customer site documentation.

— Ordered hardware is available.

— After unpacking, the AIR must be powered up within 48 hours, if high humidity or cold temperatures apply. Ensure that the AIR is not left in an outdoor environment for more than one week before taking it into service.

— The power supply (−48 V DC) for the AIR is installed and tested at the site.

— Site grounding is installed and tested.

Do!

Products not connected to ground risk damage from overvoltage or overcurrent. Always ground products according to instructions.

— The following equipment must be installed and tested:

- A main switch
- A circuit breaker or an external fuse

1.1 Safety and Operation Requirements

> Precautions from an environmental and mechanical perspective for a safe and approved installation of the product are described.

Adhere to the following for safety and operation reasons: The mechanical design of the AIR is based on environmental conditions that are equal to or exceeding class 4.1 as specified in EN 300 019-1-4 and GR-3178-CORE and thereby respects the static mechanical load imposed on an AIR by wind at maximum velocity. Wind loads in this document are calculated with reference to wind pressure. For more accurate results, the specific terrain information for relevant sites and geographical area where the AIR will be installed must be carefully analyzed, considered, and calculated according to EN 1991-1-4.

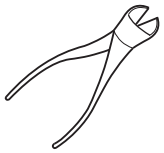
Pole clamps, brackets, mounting accessories and other installation material or equipment specified by Ericsson in the AIR product information documentation must be used and Ericsson installation instructions be complied with. In addition, it must be observed that specific environmental conditions that the AIR becomes exposed to, such as icing, heat, dust, dynamic stress (for example, strain caused by oscillating support structures) or other environmental conditions that exceed or otherwise deviate from the Antenna Integrated Radio Unit Description, can result in the breakage of an AIR or its mounting accessories and even cause the AIR to fall to the ground.

These facts, information, and circumstances must be considered and properly taken into account during the site planning process and adhered to for installation and operation of the AIR. Ericsson expressly disclaims any responsibility or liability arising out of failures in this regard.

2 Unpack and Check Materials

Environmental considerations, steps for inspection and unpacking, information about the warranty seal and safe handling of the unit are provided.

Required Tools and Equipment

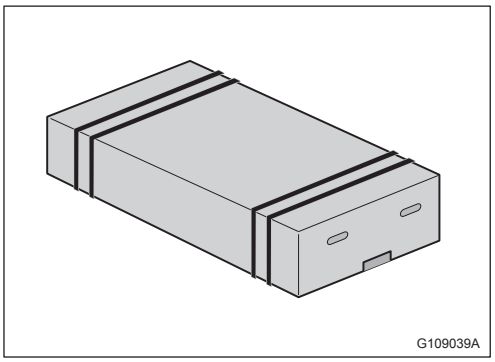
Tool	Product Name
	Side cutters
	Knife

Prerequisites

— The packed radio must be kept in weather-protected environment. The radio, once unpacked, must be installed according to the instructions, even if it is not powered up immediately.

— The radio must be powered up within 48 hours in outdoor environment.

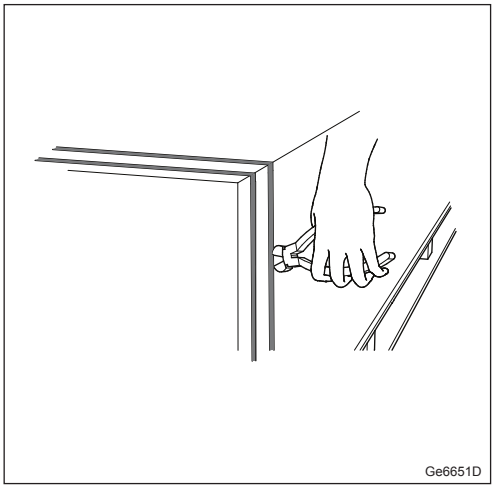
Steps

1. 

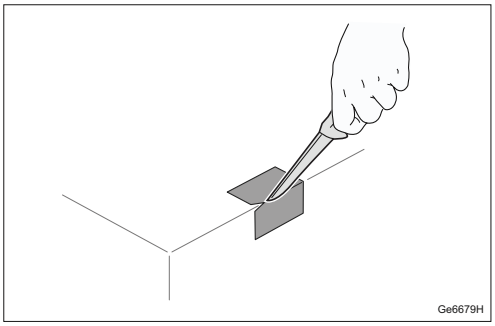
Examine the box and packing material for damage. If there are holes in the box, immediately file a complaint with the transport company. Take a photo of the damages to facilitate the claims process.

Note:

Unpack the unit only at the site, shortly before installation.

2. 

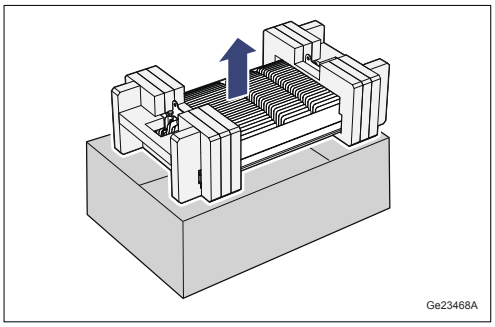
Remove the straps from the box.

3. 

Remove the tape from the box.

Note:

Ensure not to damage the items in the box.

4. 

Open the box. Take out the AIR along with the plastic foam.

Note: Handle the parts with care, and protect the parts from water and dirt.

5.

Verify that the items delivered correspond to the packing list and that they are not damaged.

If any of the items are damaged, see Handling Faulty Equipment.

6. Save the box and packing material for any return of units. For more information about packing and returning units, see Handling Faulty Equipment.

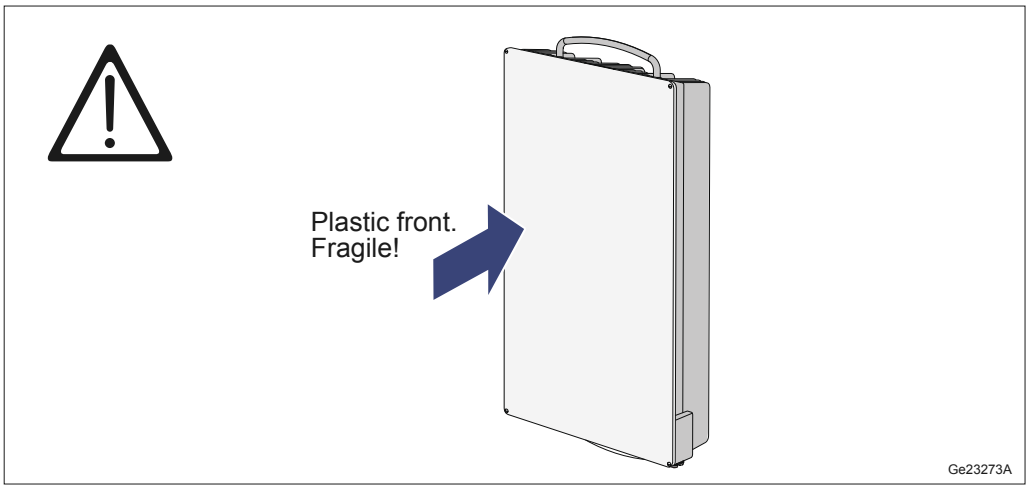
2.1 Handle AIR Unit

Information about warranty seal, lifting heavy equipment, and handling the AIR unit to avoid damage.

— The product is equipped with a warranty seal sticker.

Note: Seals that have been implemented by Ericsson must not be broken or removed, as it otherwise voids warranty.

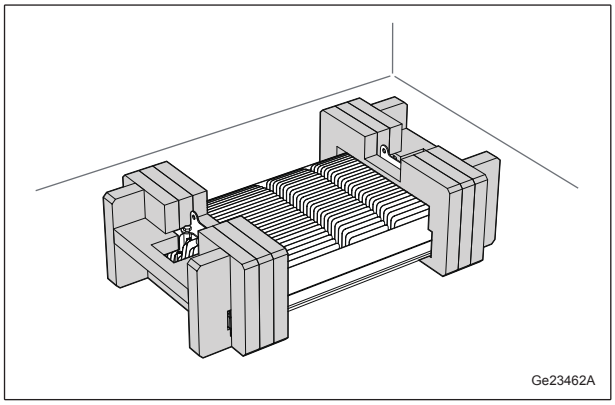
— The AIR has a plastic front. Be careful when handling it.



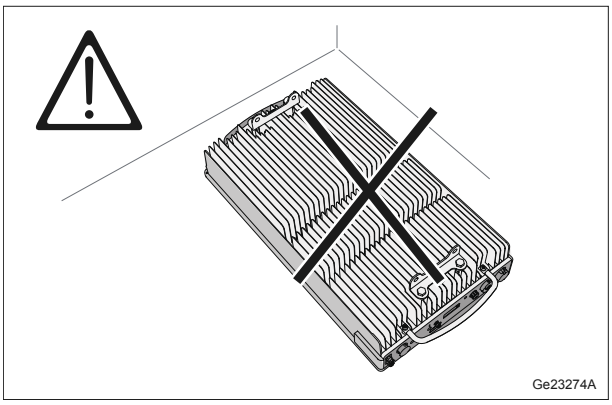
— Always try to keep the AIR in the packing material when on the ground to avoid damaging the unit.

— When lifting heavy equipment, make sure to comply with local requirements and regulations regarding maximum weight per person to be lifted.

— To protect the radome, rest the AIR on the plastic foam.

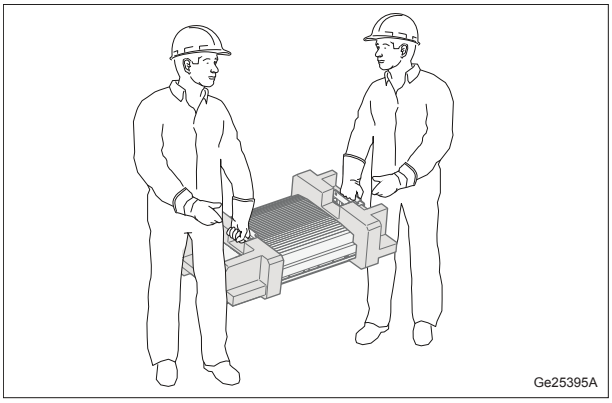


— Do not rest the AIR on the radome without the plastic foam.

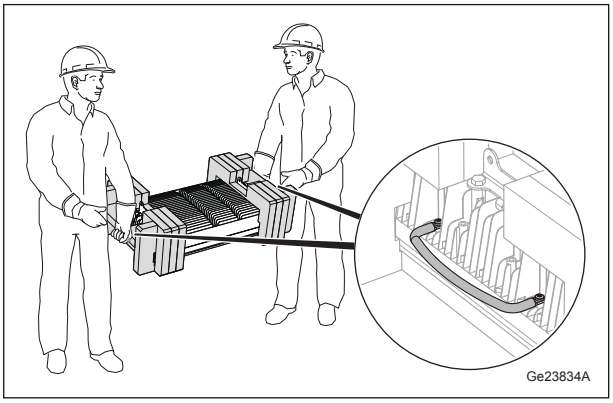


— Lift the AIR by hand.

AIR 6419 B41:

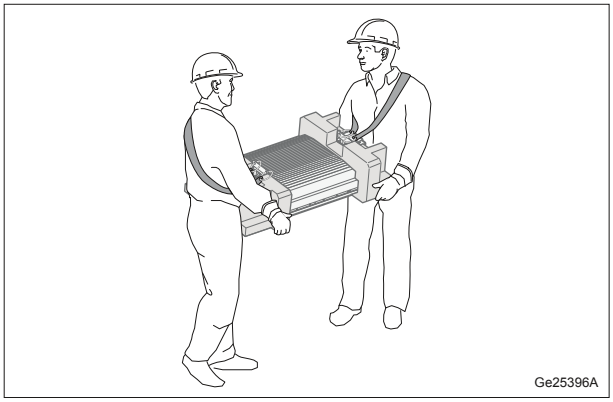


All other frequency variants:

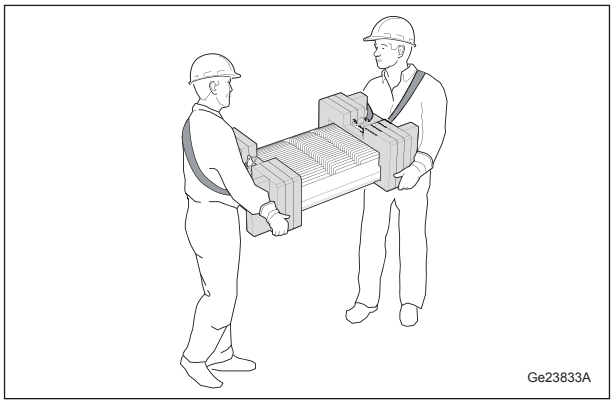


— Lift the AIR with belt.

AIR 6419 B41:



All other frequency variants:



3 Pole Installation

3.1 Prepare AIR Unit for Single Unit Pole Installation

> Before installing the AIR unit, the mounting brackets need to be assembled and mounted to the AIR unit. An installation with down tilt is shown here. The procedure for installation with up tilt or without any tilt is similar.

Required Tools and Equipment

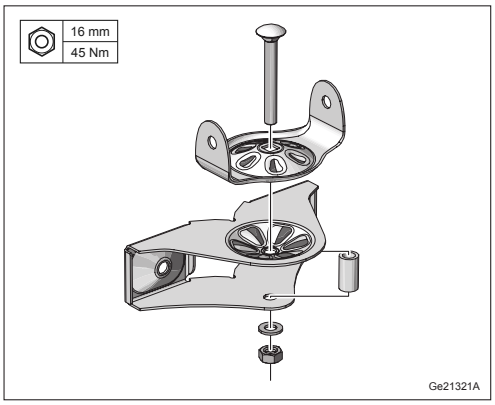
Tool	Product Name
	Torque wrench — 8–52 Nm
	Open-ended wrench head — 16 mm(1)

(1) Two 16 mm wrenches are needed to install the product

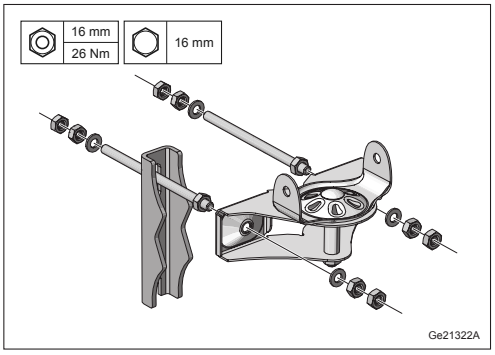
Mounting Kit	Product Name
Swivel mounting kit	AIR Light, wall and pole mount bracket no tilt with azimuth
Tilt and swivel mounting kit	AIR Light, wall and pole mount bracket with tilt and azimuth

Steps

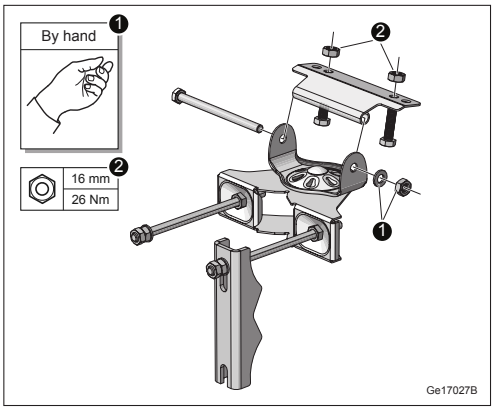
If the brackets are preassembled, skip Step 1 through Step 5.

1. 

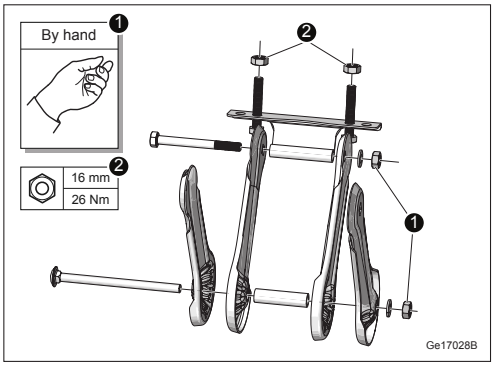
   Install the azimuth joint plate to each of the two azimuth clamps.

2. 

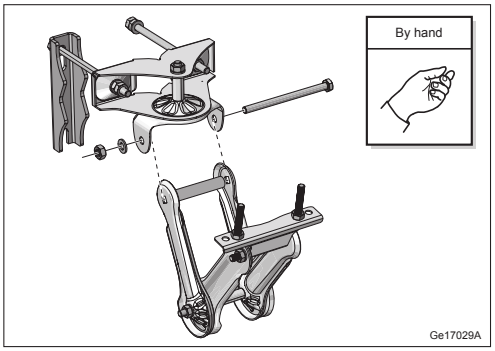
Install the clamp to each of the two azimuth clamp assemblies.

3. 

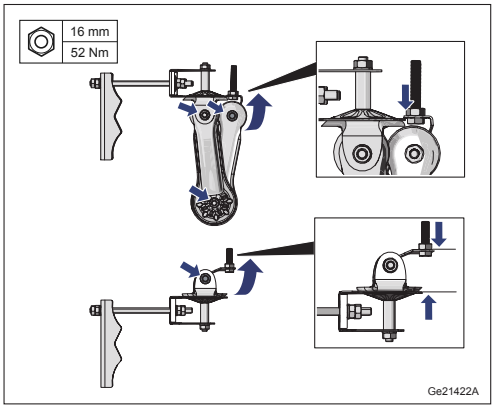
Install the bottom tilt joint to one of the azimuth clamp assemblies.

4. 

Assemble the tilt arm.

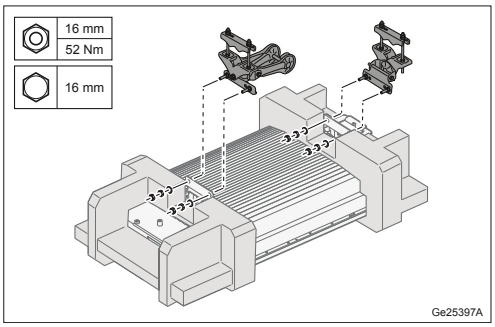
5. 

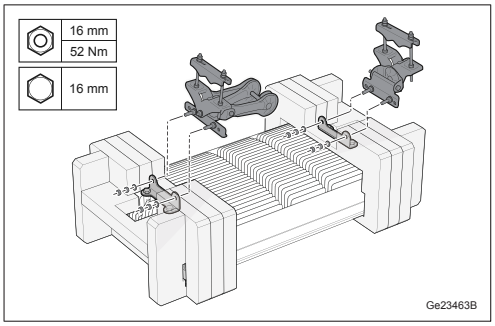
Install the tilt arm to the other azimuth clamp assembly.

6. 

Adjust the tilt and swivel angles to 0°.

Tighten the tilt bolts.

7. 



Install the clamp assemblies onto the AIR.

Break the plastic foam bars that may block the installation of the mounting kit, if applicable.

The first illustration is valid for AIR 6419 B41.

The second illustration is valid for all other frequency variants.

8. Make sure that the bracket hinges are fully closed and that all the nuts are pretightened enough to avoid movement during hoisting.

3.2 Install SFP Module

> SFP modules must be installed to enable optical communication for the unit.

Required Tools and Equipment

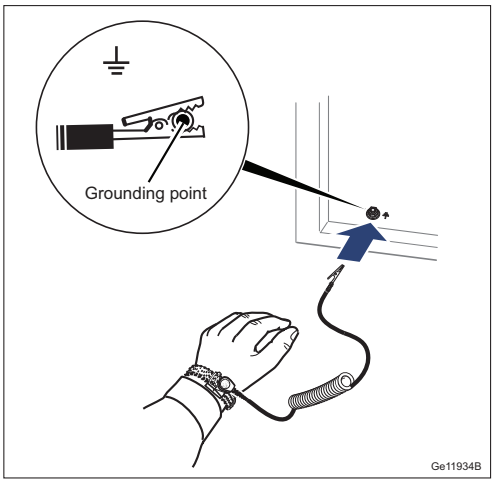
Tool	Product Name
	ESD wrist strap

Prerequisites

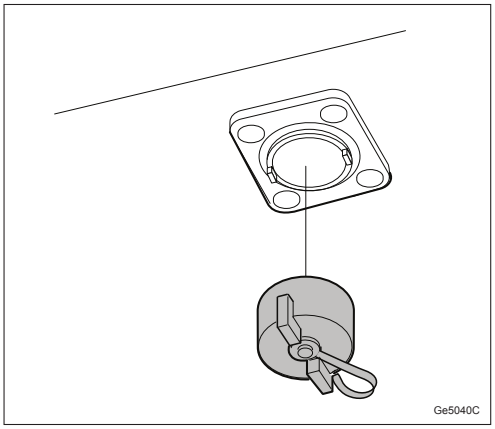
Only use SFP modules approved and supplied by Ericsson.

Be careful when handling SFP modules. For information on how to handle SFP modules, see Handling SFP Modules and Optical Cables.

Steps

1. 

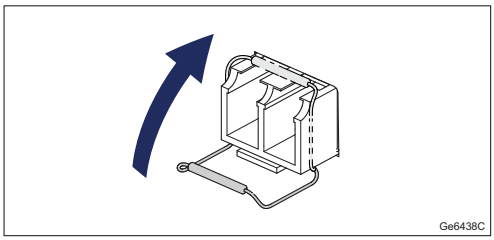
   Connect the ESD wrist strap to the grounding point.

2. 

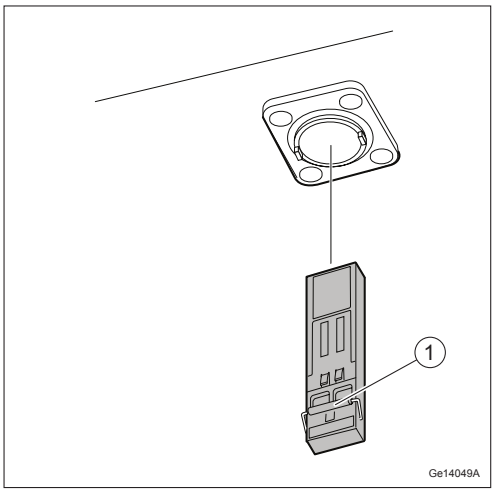
Remove the cap.

Note:

If the optical interface uses the red throw type cap without locking feature, that optical interface must be used first because its cap is not waterproof.

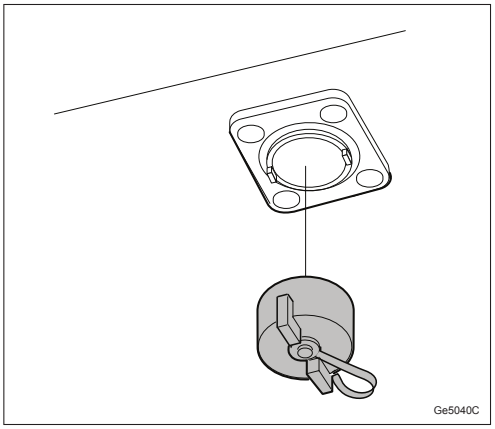
3. 

If applicable, close the latch on the SFP module.

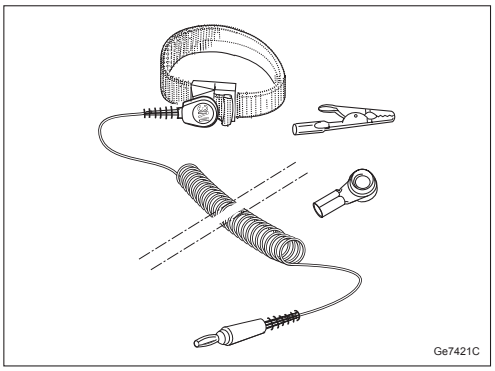
4. 

Install the SFP module in the SFP slot.

Note: To avoid equipment damage, make sure to insert the SFP module correctly oriented. The bale (1) must be oriented in the same direction as the radome.

5. 

Refit the cap.

6. 

Disconnect the ESD wrist strap.

3.3 Install RAE Unit (Optional)

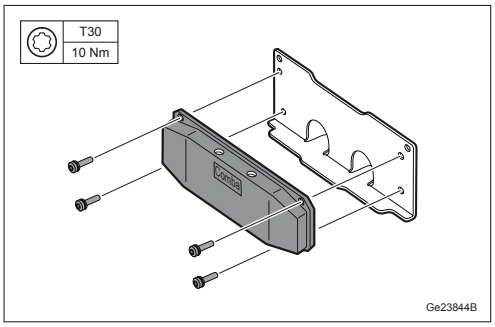
Required Tools and Equipment

Tool	Product Name
	Torque wrench — 10 Nm
	Bits — Torx T30

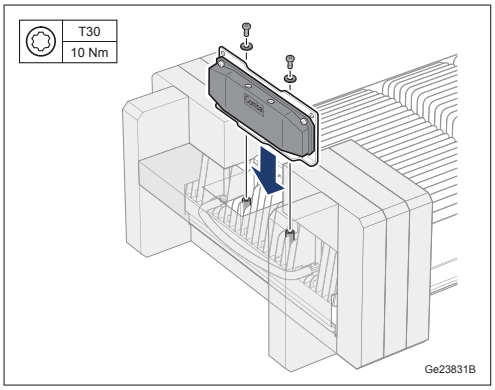
Prerequisites

— The temperature must be between 0°C to +40°C when installing RAE.

Steps

1. 

Assemble the RAE unit to the RAE holder. Attach the washers and tighten the bolts. Hold the nuts and tighten the bolts.

2. 

Mount the RAE unit on top of the unit by tightening the two bolts.

3.4 Assemble DC Power Connector to Shielded Cable

> Before assembling the DC power connector, the cable must be stripped according to specification and the shield must be connected correctly.

Stop!

Improper installation or modification of AC or DC powered equipment can seriously damage the equipment. Only a qualified and authorized electrician is permitted to install or modify electrical installations.

Required Tools and Equipment

See required tools in the assembly instruction delivered with the DC power connector.

Prerequisites

— Supplier assembly instruction

— One of the following power connector models must be used.

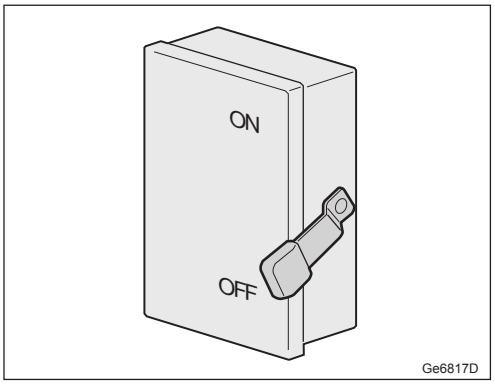
— For 2-wire (DC-C), 7F2RS104B-0M16T1 (RNT 447 37/01).

— For 3-wire (DC-I), 7F2RS103B-0M16T1 (RNT 447 36/01)

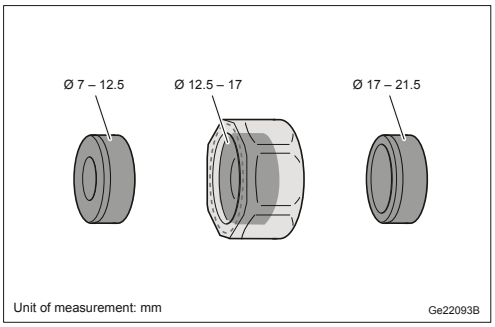
Model number is printed on the power connector.

Note: This instruction does not replace the supplier assembly instruction delivered with the DC power connector. Always read the supplier assembly instruction.

Steps

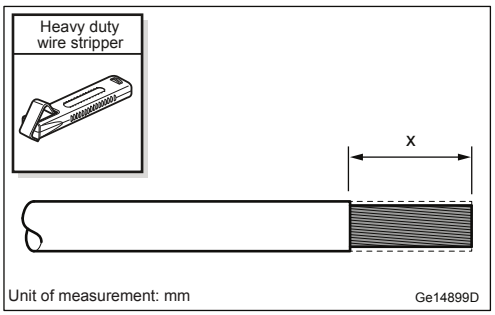
1. 

Make sure that the external power is switched off before you connect the cable.

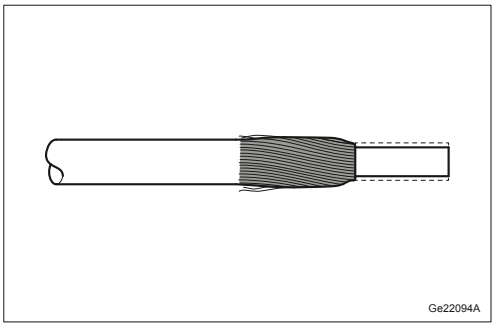
2. 

Identify the outer diameter of the power cable.

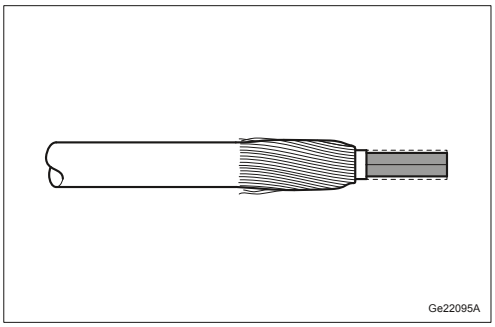
If necessary, replace the gasket.

3. 

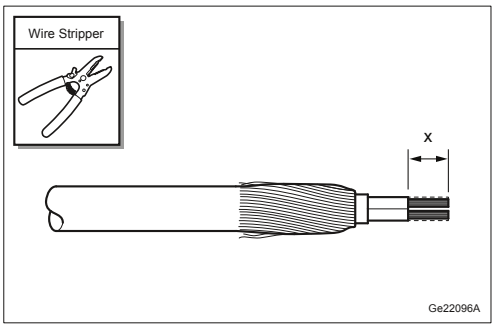
Strip the cable outer jacket according to the instructions delivered with the DC power connector.

4. 

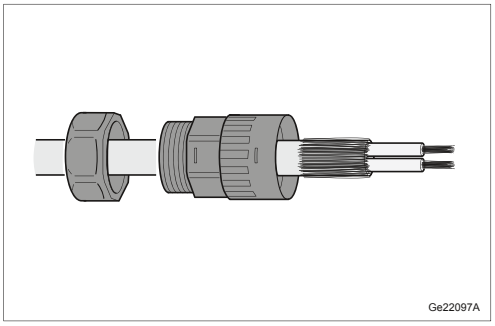
Fold back the shielding braid on the outer jacket.

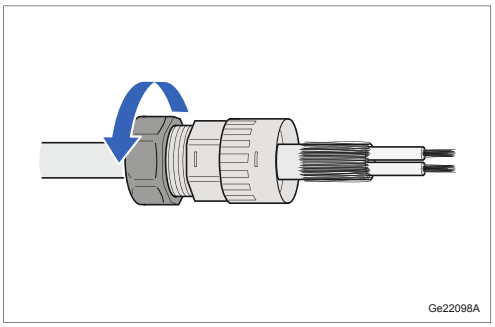
5. 

Remove the second insulation.

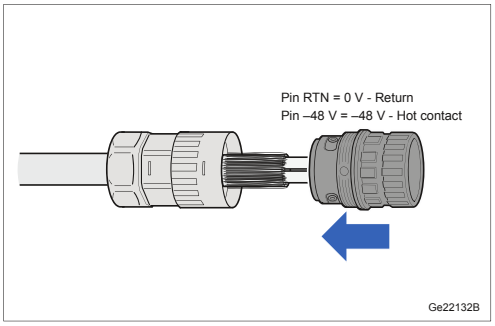
6. 

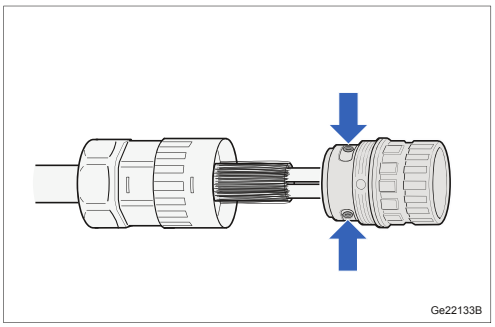
Strip the inner jacket of the conductor according to the instructions delivered with the DC power connector.

7. 



Secure the connector body on the cable according to the instructions delivered with the DC power connector.

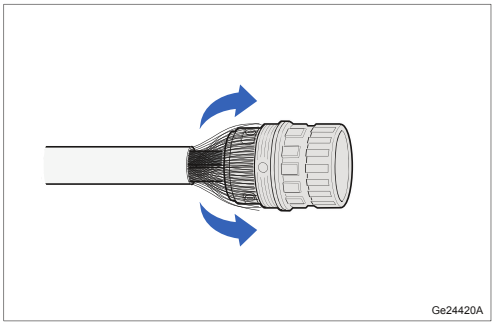
8. 



Assemble the connector according to the instructions delivered with the DC power connector.

Note:

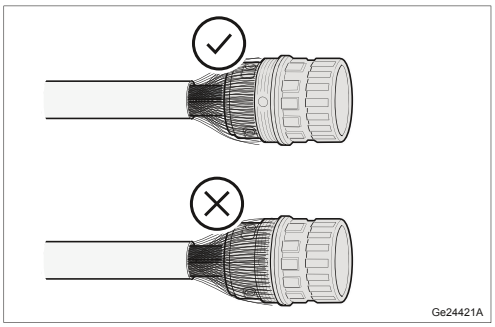
The wires of the power cable can have different colors. Make sure that the wires are connected to the correct polarities.

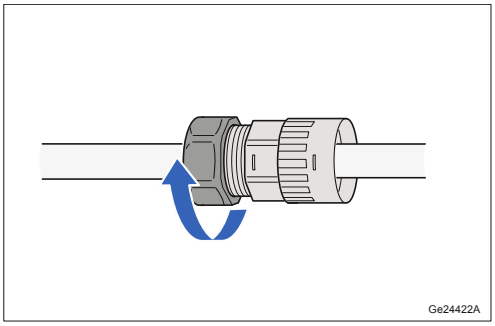
9. 

Fold the shielding braid on the DC power connector according to the instructions delivered with the DC power connector.

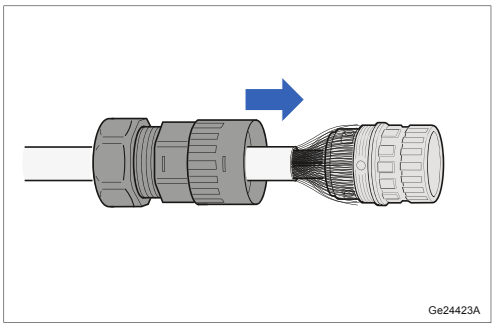
Note:

Make sure that there is no braid over the thread on the connector.

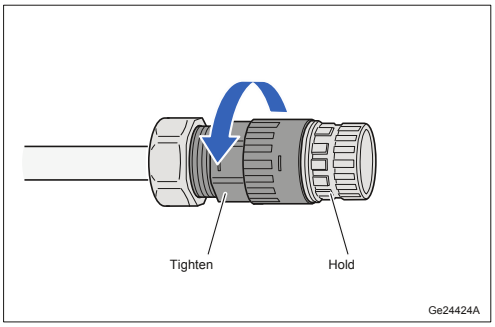


10. 

Unscrew the nut so that the connection body can be moved on the cable.

11. 

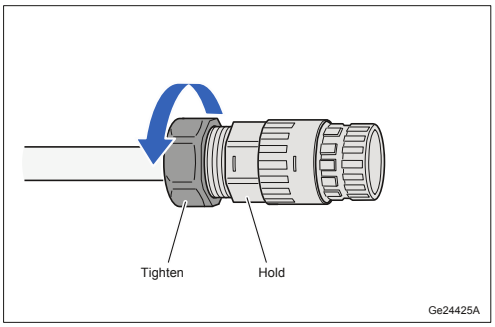
Move the connection body to the connector.

12. 

Assemble the connection body to the connector. Tighten the connection body according to the instructions delivered with the DC power connector.

Note:

Use supplier-recommended wrenches and torque to tighten.

13. 

Tighten the nut to the connection body according to the instructions delivered with the DC power connector.

Note:

Use supplier-recommended wrenches and torque to tighten.

Do!

Products not connected to ground risk damage from overvoltage or overcurrent. Always ground products according to instructions.

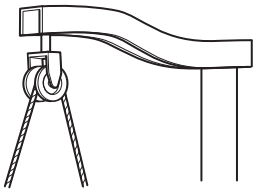
Make sure that the cable shield is connected correctly to secure lightning protection and grounding in the DC power connector.

3.5 Install AIR Unit on a Pole

With the attached mounting brackets and installed SFP modules, the unit can be hoisted and installed on the pole. An installation with tilt bracket is

shown here. The procedure for installation with bracket without tilt is similar.

Required Tools and Equipment

Tool	Product Name
	Lifting device
	Helmet
	Lubricant
	Torque wrench — 8–52 Nm
	Open-ended wrench head — 16 mm(1)

(1) Two 16 mm wrenches are needed to install the product

Prerequisites

Warning!

When working at height, there is a risk of falling objects. Falling objects can cause serious injury or even be fatal. Always wear a helmet and avoid standing in an area where objects can fall.

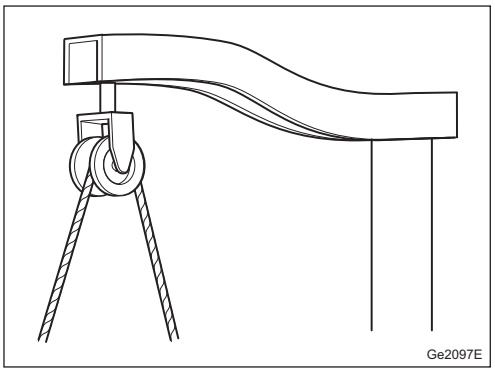
Note: Ensure that bolts and nuts are lubricated.

Do not use lubricants containing graphite since they cause corrosion and can damage cable insulation.

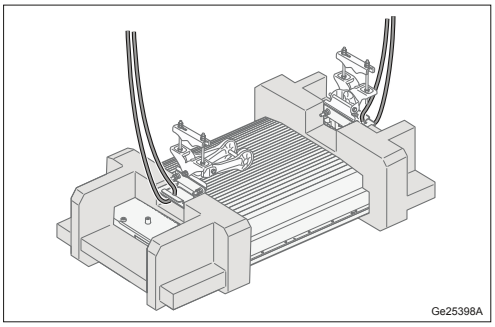
— Be careful not to damage the unit against the pole, mast, or wall. Always secure the unit properly.

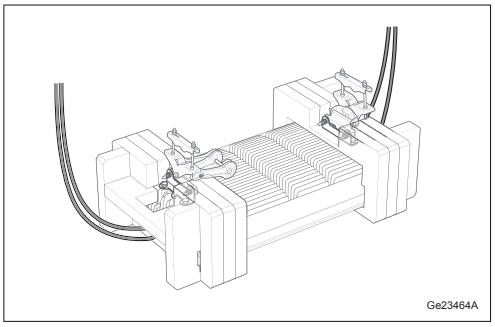
— When hoisting, make sure to comply with local requirements and regulations regarding hoisting equipment.

Steps

1. 

Attach a lifting device onto the top of the pole or wall where the AIR is to be attached.

2. 

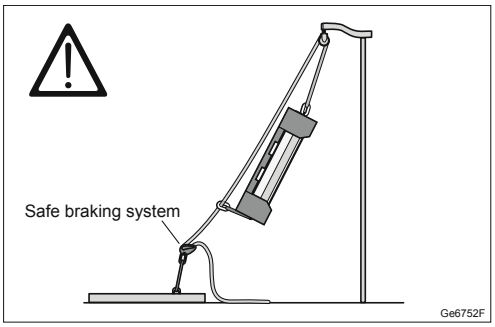


Pull a rope through the block on the lifting device and attach the end of the rope closest to the pole or wall to the lifting eye on the top part of the AIR. Attach it either directly or through a snap link. With a snap link, attach a rope to the lifting eye on the lower part of the AIR.

To prevent movement during hoisting, tighten the tilt bracket bolts.

The first illustration is valid for AIR 6419 B41.

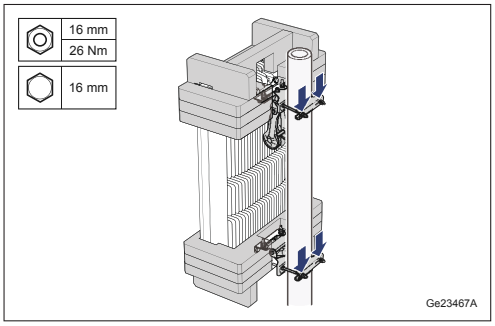
The second illustration is valid for all other frequency variants.

3. 

Hoist the AIR with the plastic foam up to the desired position on the pole or wall. Use the lower rope to prevent the AIR from hitting the pole or wall, or other units.

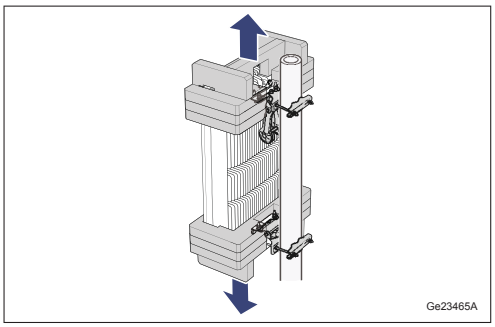
Note:

Use a safe braking system.

4. 

Close the pole clamps around the pole, and finger tighten the nuts.

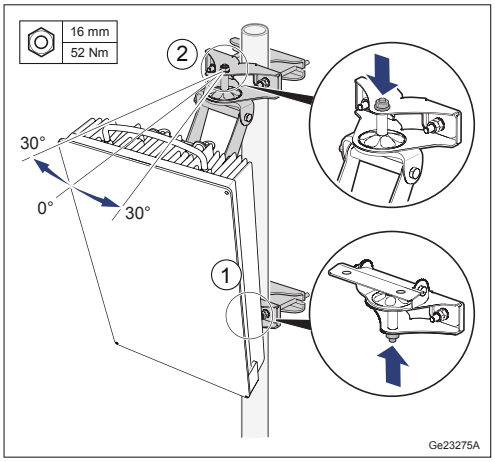
Adjust the pole clamp so that the clamp bolts run parallel on each side of the pole. Tighten the outer M10 nuts.

5. 

Remove the lifting strap, the plastic foam, and the bubble film from the AIR. Save the packing material for any return of units.

Note:

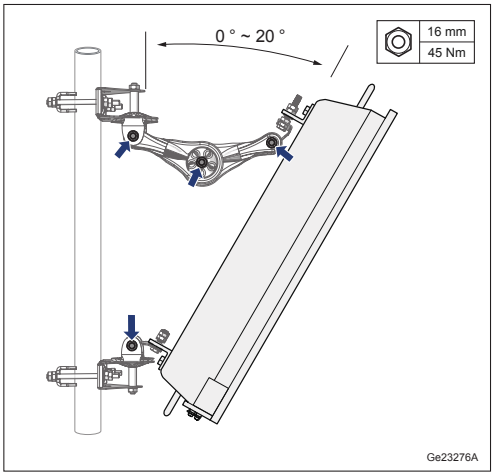
Before removing the plastic foam and bubble film, make sure that they are tethered at all times.

6. 

To adjust the left and right angle, loosen the bolts and set the angle on upper and lower bracket. Do not remove any nuts.

After adjusting the angle, retighten the bolts.

7.



For installation with tilt bracket:

To adjust the down or up tilt angle, loosen the bolts and set the angle. Do not remove any nuts.

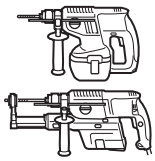
After adjusting the angle, retighten the bolts.

4 Wall Installation

4.1 Prepare AIR Unit for Single Unit Wall Installation

> Before installing the AIR unit, the mounting brackets need to be assembled and mounted to the AIR unit. An installation with down tilt is shown here. The procedure for installation with up tilt or without any tilt is similar.

Required Tools and Equipment

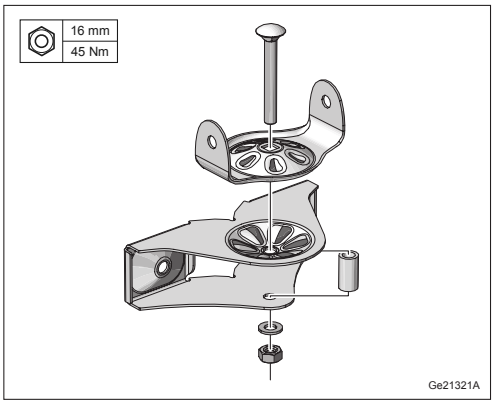
Tool	Product Name
	Torque wrench — 8–52 Nm
	Open-ended wrench head — 16 mm(1)
	Ear protection
	Eye protection
	Hammer drill with depth gauge
	Measure
	Pen
	Spirit level

(1) Two 16 mm wrenches are needed to install the product

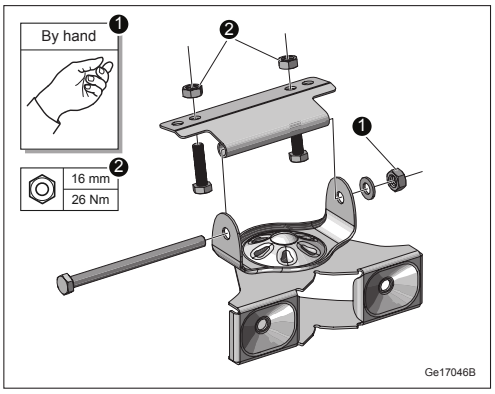
Mounting Kit	Product Name
Swivel mounting kit	Wall and Pole clamp no tilt with azimuth
Tilt and swivel mounting kit	Wall and Pole clamp with tilt and azimuth

Steps

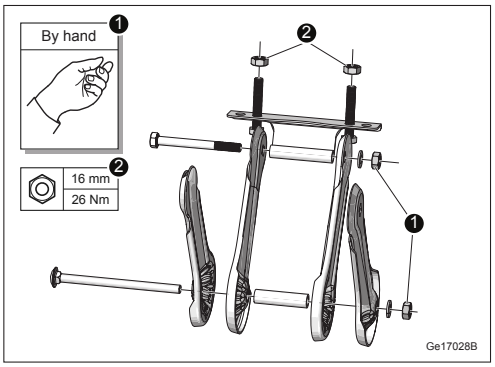
If the brackets are preassembled, skip Step 1 through Step 4.

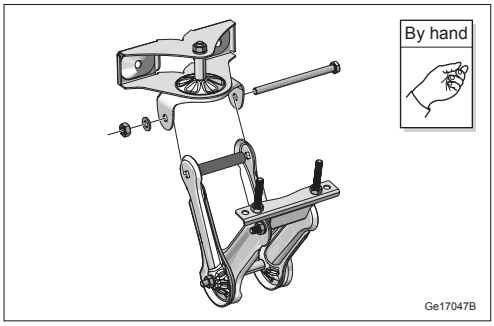
1. 

Install the azimuth joint plate to each of the two azimuth clamps.

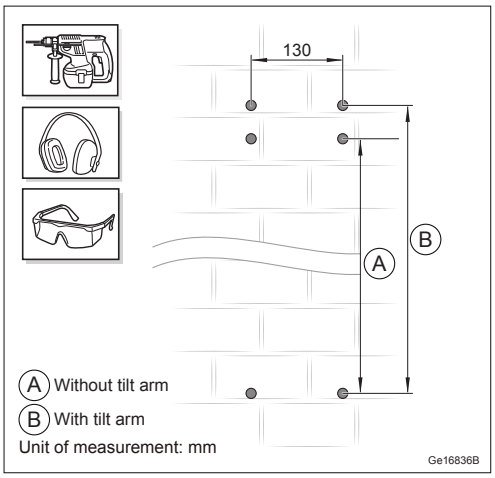
2. 

Install the bottom tilt joint to one of the azimuth clamp assemblies.

3. 

4. 

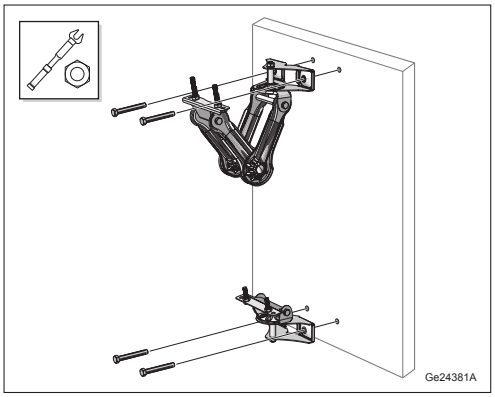
Install the tilt arm to the other azimuth clamp assembly.

5. 

Drill four holes in the wall using the illustration for hole positions, and matching the bolts and plugs specified by the site engineer.

Fit a depth restrictor to the hammer drill to regulate the depth of each hole. Wear protective equipment when drilling.

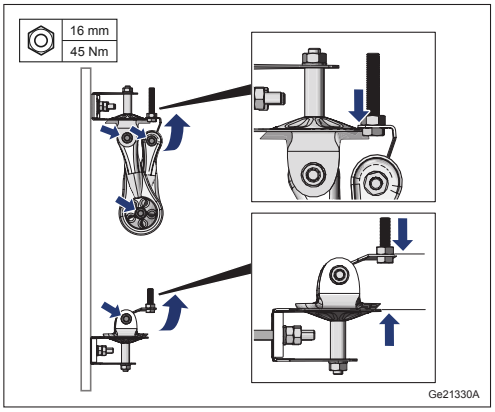
A=581 mm, B=687 mm.

6. 

Install the tilt and no tilt clamp assembly onto the wall.

Tighten the bolts according to recommendations given by site engineers.

Bolts, plugs, or expander bolts for securing wall brackets to the wall are not delivered with the unit. The types of bolts and plugs must be specified by the site engineers.

7. 

Adjust the tilt and swivel angles to 0°.

Tighten the tilt bolts.

8. Make sure that the bracket hinges are fully closed and that all the nuts are pretightened enough to avoid movement during hoisting.

4.2 Install SFP Module

> SFP modules must be installed to enable optical communication for the unit.

Required Tools and Equipment

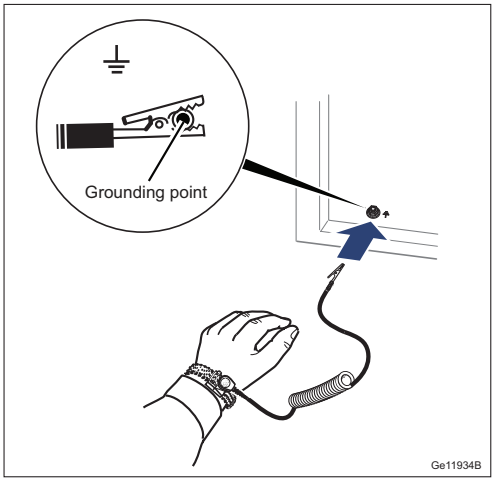
Tool	Product Name
	ESD wrist strap

Prerequisites

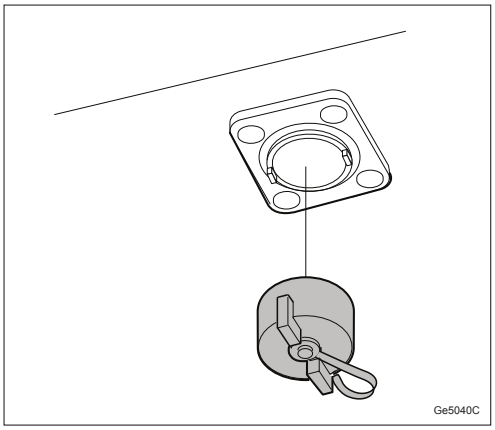
Only use SFP modules approved and supplied by Ericsson.

Be careful when handling SFP modules. For information on how to handle SFP modules, see Handling SFP Modules and Optical Cables.

Steps

1. 

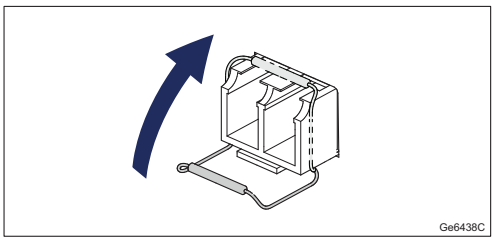
Connect the ESD wrist strap to the grounding point.

2. 

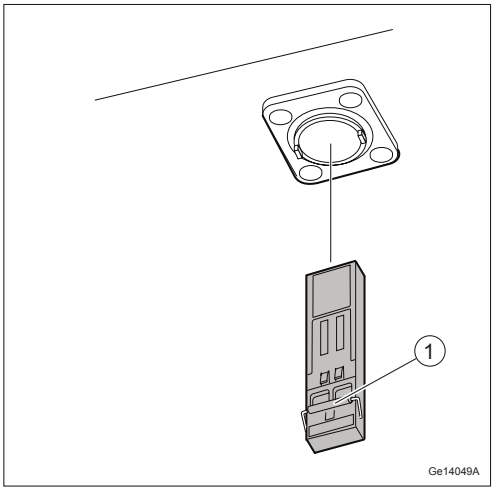
Remove the cap.

Note:

If the optical interface uses the red throw type cap without locking feature, that optical interface must be used first because its cap is not waterproof.

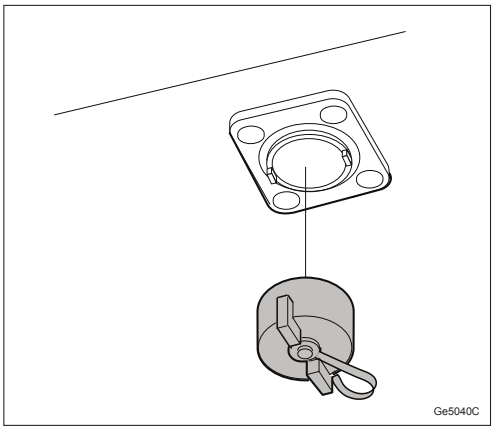
3. 

If applicable, close the latch on the SFP module.

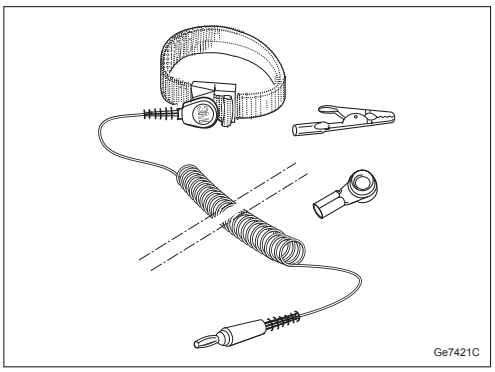
4. 

Install the SFP module in the SFP slot.

Note: To avoid equipment damage, make sure to insert the SFP module correctly oriented. The bale (1) must be oriented in the same direction as the radome.

5. 

Refit the cap.

6. 

Disconnect the ESD wrist strap.

4.3 Install RAE Unit (Optional)

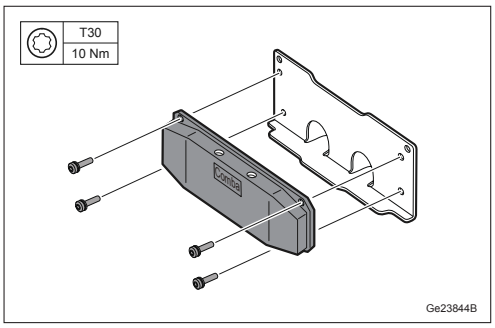
Required Tools and Equipment

Tool	Product Name
	Torque wrench — 10 Nm
	Bits — Torx T30

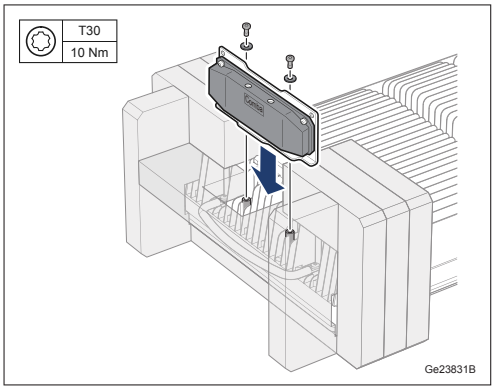
Prerequisites

— The temperature must be between 0°C to +40°C when installing RAE.

Steps

1. 

Assemble the RAE unit to the RAE holder. Attach the washers and tighten the bolts. Hold the nuts and tighten the bolts.

2. 

Mount the RAE unit on top of the unit by tightening the two bolts.

4.4 Assemble DC Power Connector to Shielded Cable

> Before assembling the DC power connector, the cable must be stripped according to specification and the shield must be connected correctly.

Stop!

Improper installation or modification of AC or DC powered equipment can seriously damage the equipment. Only a qualified and authorized electrician is permitted to install or modify electrical installations.

Required Tools and Equipment

See required tools in the assembly instruction delivered with the DC power connector.

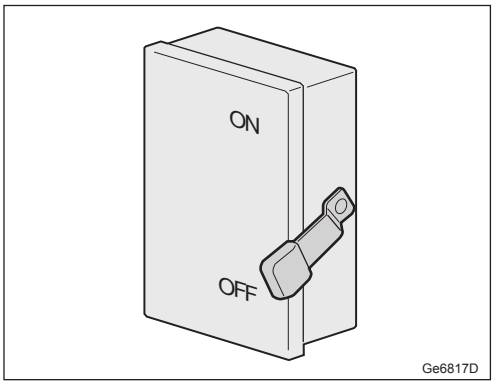
Prerequisites

- Supplier assembly instruction
- One of the following power connector models must be used.
  - For 2-wire (DC-C), 7F2RS104B-0M16T1 (RNT 447 37/01).
  - For 3-wire (DC-I), 7F2RS103B-0M16T1 (RNT 447 36/01)

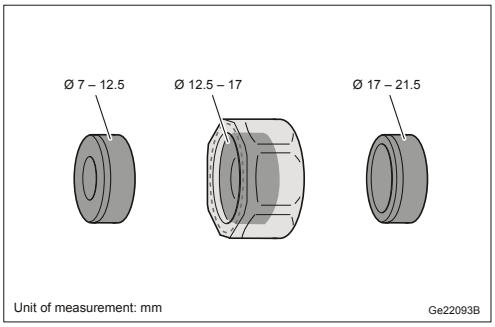
  Model number is printed on the power connector.

Note: This instruction does not replace the supplier assembly instruction delivered with the DC power connector. Always read the supplier assembly instruction.

Steps

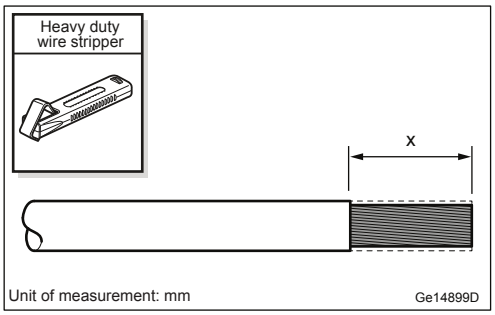
1. 

   Make sure that the external power is switched off before you connect the cable.

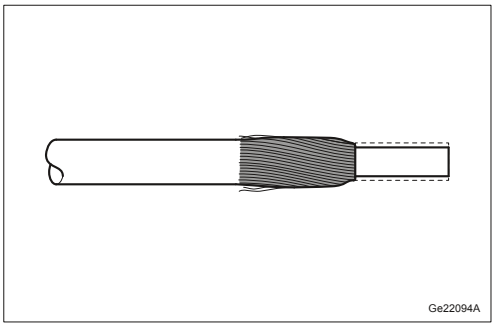
2. 

Identify the outer diameter of the power cable.

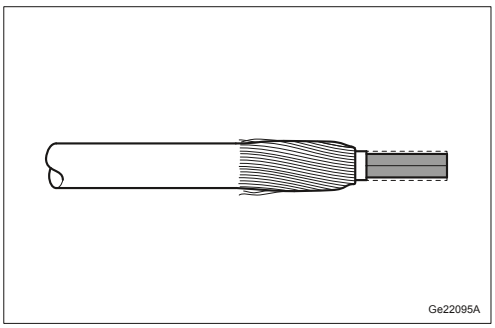
If necessary, replace the gasket.

3. 

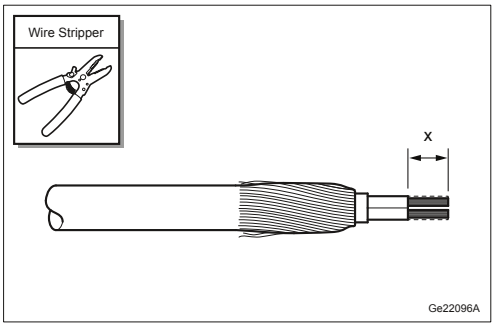
Strip the cable outer jacket according to the instructions delivered with the DC power connector.

4. 

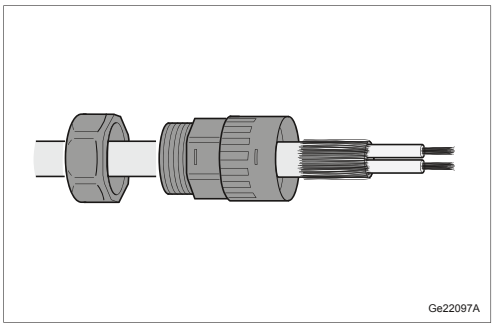
Fold back the shielding braid on the outer jacket.

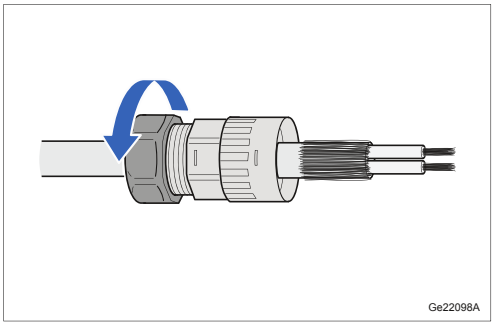
5. 

Remove the second insulation.

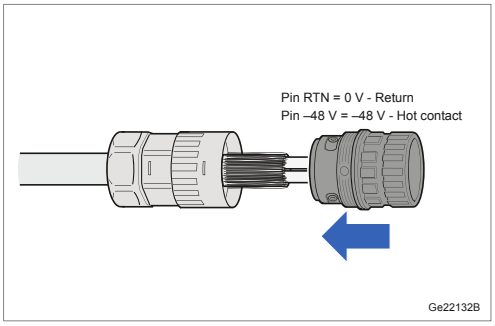
6. 

Strip the inner jacket of the conductor according to the instructions delivered with the DC power connector.

7. 



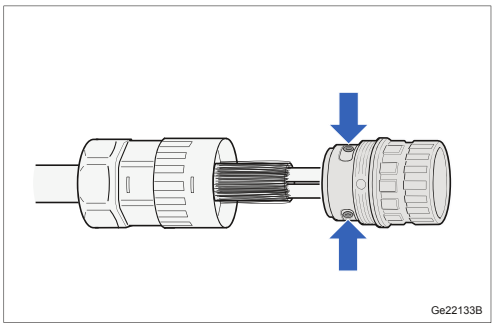
Secure the connector body on the cable according to the instructions delivered with the DC power connector.

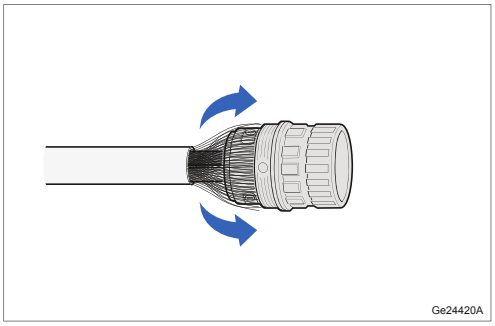
8. 

Assemble the connector according to the instructions delivered with the DC power connector.

Note:

The wires of the power cable can have different colors. Make sure that the wires are connected to the correct polarities.

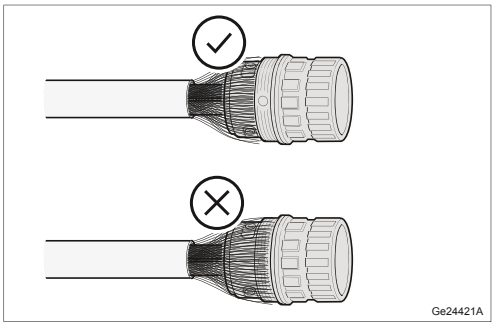


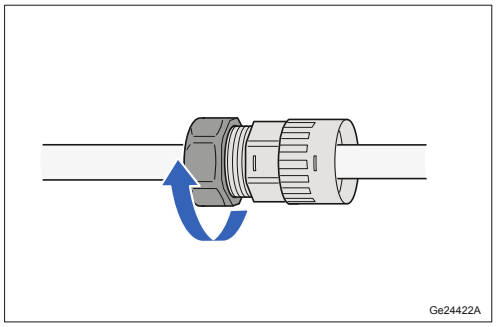
9. 

Fold the shielding braid on the DC power connector according to the instructions delivered with the DC power connector.

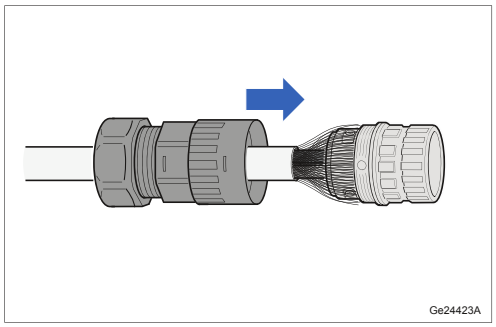
Note:

Make sure that there is no braid over the thread on the connector.

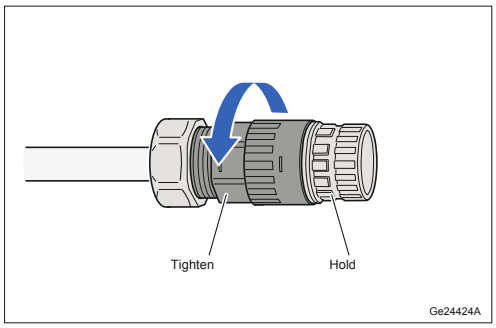


10. 

Unscrew the nut so that the connection body can be moved on the cable.

11. 

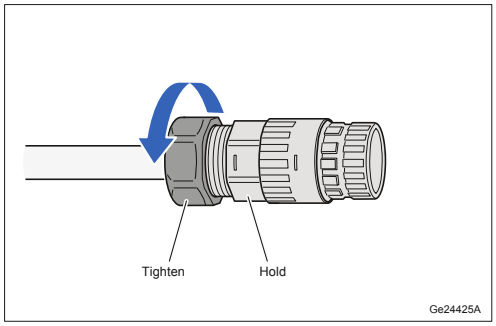
Move the connection body to the connector.

12. 

Assemble the connection body to the connector. Tighten the connection body according to the instructions delivered with the DC power connector.

Note:

Use supplier-recommended wrenches and torque to tighten.

13. 

Tighten the nut to the connection body according to the instructions delivered with the DC power connector.

Note:

Use supplier-recommended wrenches and torque to tighten.

Do!

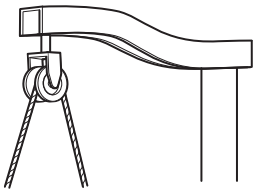
Products not connected to ground risk damage from overvoltage or overcurrent. Always ground products according to instructions.

Make sure that the cable shield is connected correctly to secure lightning protection and grounding in the DC power connector.

4.5 Install AIR Unit on a Wall

With the attached mounting brackets and installed SFP modules, the unit can be hoisted and installed on the wall. An installation with tilt bracket is shown here. The procedure for installation with bracket without tilt is similar.

Required Tools and Equipment

Tool	Product Name
	Lifting device
	Helmet
	Lubricant
	Torque wrench — 8–52 Nm
	Open-ended wrench head — 16 mm(1)

(1) Two 16 mm wrenches are needed to install the product

Prerequisites

Warning!

When working at height, there is a risk of falling objects. Falling objects can cause serious injury or even be fatal. Always wear a helmet and avoid standing in an area where objects can fall.

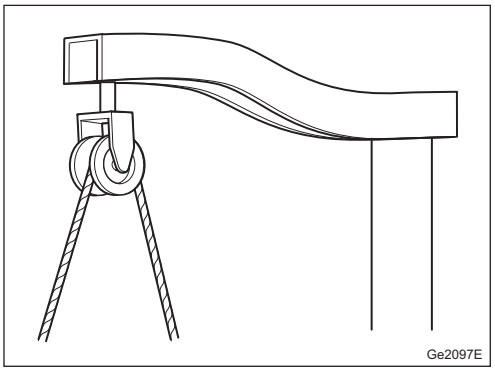
Note: Ensure that bolts and nuts are lubricated.

Do not use lubricants containing graphite since they cause corrosion and can damage cable insulation.

— Be careful not to damage the unit against the pole, mast, or wall. Always secure the unit properly.

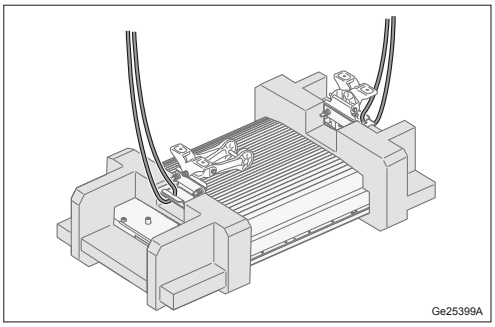
— When hoisting, make sure to comply with local requirements and regulations regarding hoisting equipment.

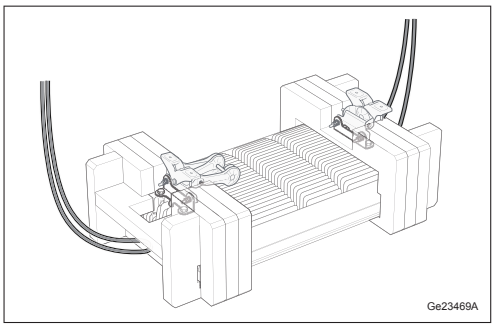
Steps

1. 

Attach a lifting device onto the top of the pole or wall where the AIR is to be attached.

2.





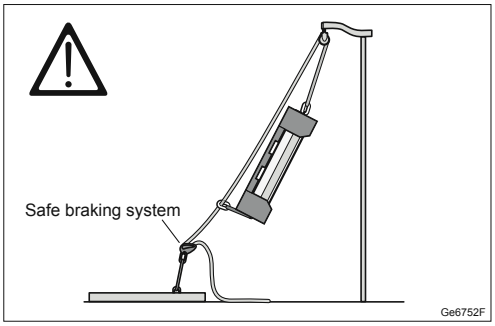
Pull a rope through the block on the lifting device and attach the end of the rope closest to the pole or wall to the lifting eye on the top part of the AIR. Attach it either directly or through a snap link. With a snap link, attach a rope to the lifting eye on the lower part of the AIR.

Break the foam bars that may block the installation of the mounting kit, if applicable.

The first illustration is valid for AIR 6419 B41.

The second illustration is valid for all other frequency variants.

3.

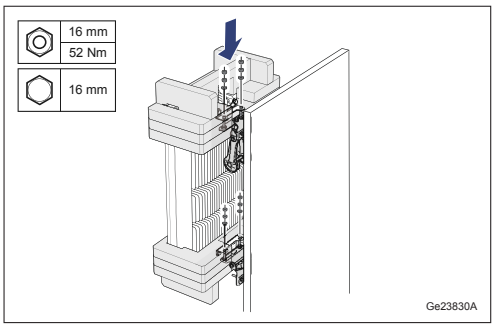


Hoist the AIR with the plastic foam up to the desired position on the pole or wall. Use the lower rope to prevent the AIR from hitting the pole or wall, or other units.

Note:

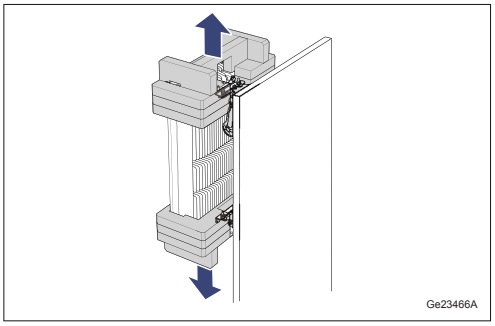
Use a safe braking system.

4.



Position the AIR onto the upper and lower wall brackets and tighten the bolts.

5.

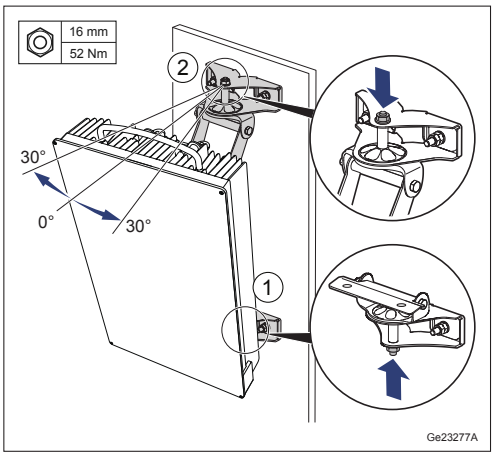


Remove the lifting strap, the plastic foam, and the bubble film from the AIR. Save the packing material for any return of units.

Note:

Before removing the plastic foam and bubble film, make sure that they are tethered at all times.

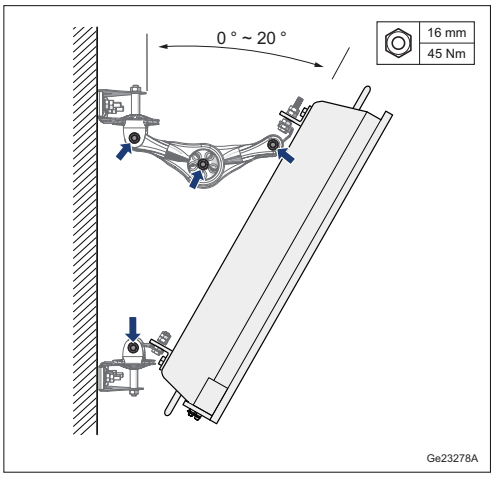
6.



To adjust the left and right angle, loosen the bolts and set the angle on upper and lower bracket. Do not remove any nuts.

After adjusting the angle, retighten the bolts.

7.



For installation with tilt bracket:

To adjust the down or up tilt angle, loosen the bolts and set the angle. Do not remove any nuts.

After adjusting the angle, retighten the bolts.

5 Connect Cables

> With the unit installed, the grounding cable, the optical cables, the EC light cable, the RAE cable, the TX monitor cable, and the power cable can be connected.

Make sure that all cable connectors are in good condition. Replace the connectors if needed.

For information on cable product numbers, see Site Installation Products Overview.

Because of space limitations, always connect the cables in the following order:

1. Connect the grounding cable.
2. Connect the optical cable.
3. Connect the EC light cable (optional).
4. Connect the RAE cable (optional).
5. Connect the external alarm cable (optional).
6. Connect the TX monitor cable (optional).
7. Connect the power cable.

5.1 Hoist Cables

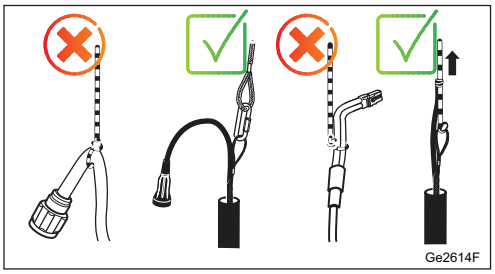
> How to hoist cables up to the AIR with a pulling wire.

Prerequisites

Optical cables are fiber optic and must be handled with care. Never touch the end-face of the optical connectors. Only remove protective covers right before installation. Observe minimum allowed bending radii. For instructions on installing the optical cable, see Installing Optical Cables for Main-Remote Solutions.

For information on how to handle optical cables, see Handling SFP Modules and Optical Cables.

Steps

1. 

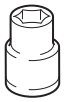
   Attach the pulling wire to the optical cable.

2. Pull the pulling wire up to the AIR. The pulling force must not exceed 400 N.
3. Clamp or strap the optical cable to the pole or wall.

5.2 Connect Grounding Cable

> The first cable to be connected is the grounding cable.

Required Tools and Equipment

Tool	Product Name
	Bits — TORX® T30(1)
	Torque wrench — 8–52 Nm
	Bit holder
	Sockets — 10 mm

(1) Use either a Torx T30 bit with a bit holder, or a 10 mm socket to connect the grounding cable.

Prerequisites

— Grounding kit

For general information about grounding principles, see Grounding Guidelines for RBS Sites.

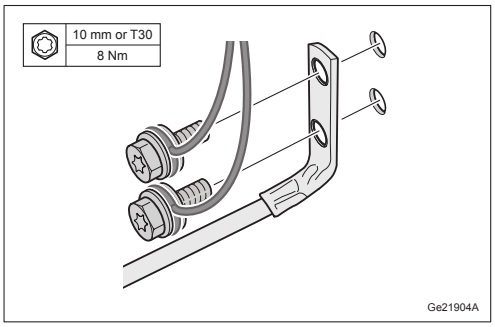
— The minimum grounding cable area is 16 mm² for copper cable, or a similar cable.

— MET installed

Do!

Products not connected to ground risk damage from overvoltage or overcurrent. Always ground products according to instructions.

Steps

1. Route the grounding cable from the MET at the site to the grounding point on the AIR.
2. 

   Unscrew the two M6 bolts from the unit and connect the grounding cable.

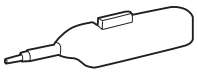
After This Task

When the installation of the AIR is finished, strap the cables to the pole or wall.

5.3 Connect Optical Cables

> After the optical cables have been hoisted up to the AIR, they can be connected to the optical connection ports.

Required Tools and Equipment

Tool	Product Name
	Cleaner for optical connections
	Ferrule end-face cleaner

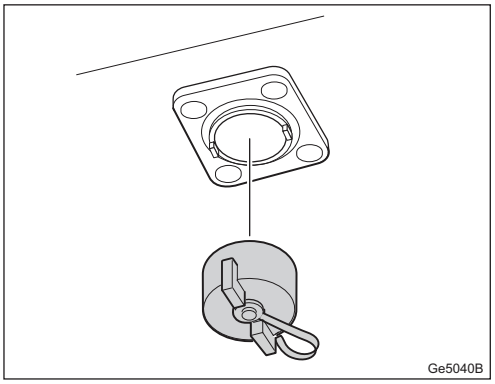
Prerequisites

Optical cables are fiber optic and must be handled with care. Never touch the end-face of the optical connectors. Only remove protective covers right before installation.

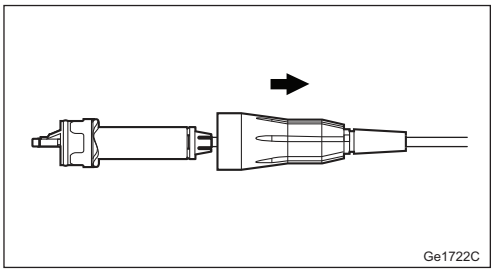
For a more detailed instruction on how to install optical cable, see Installing Optical Cables for Main-Remote Solutions.

Steps

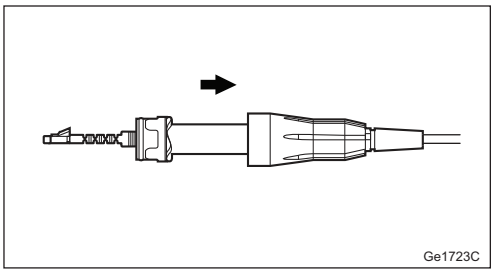
Repeat Step 1 through Step 9 for all optical connection ports.

1. 

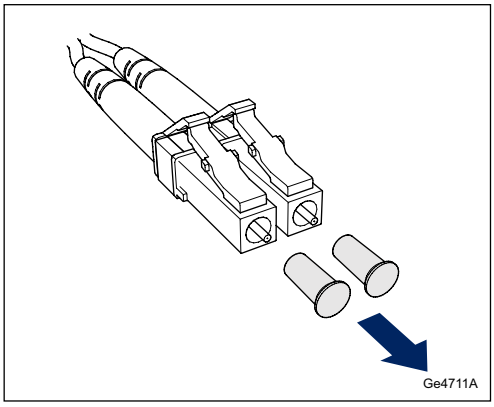
   Remove the protective cap from the optical connection port.

2. 

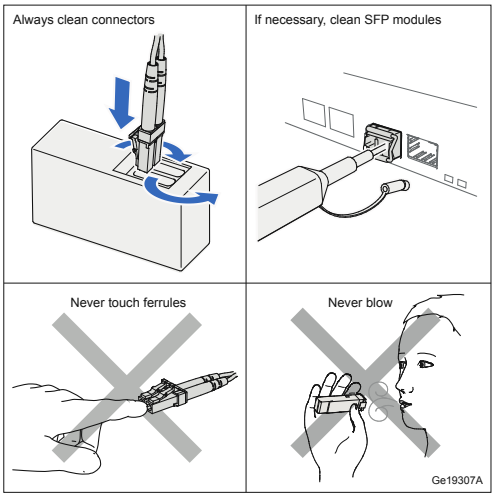
   Draw back the outer cover from the connector.

3. 

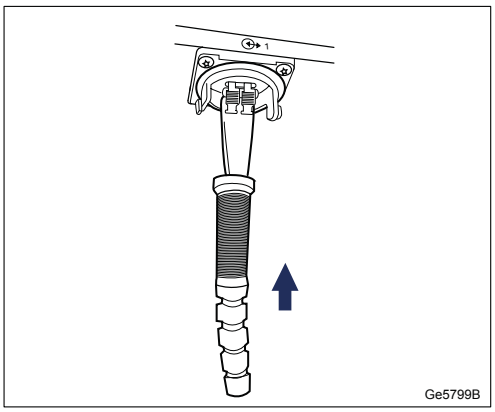
   Draw back the protective cover from the optical cable connector.

4. 

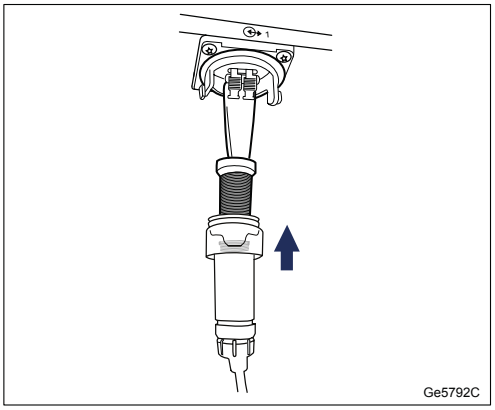
Remove the fiber protection from the fibers in the optical cable.

5. 

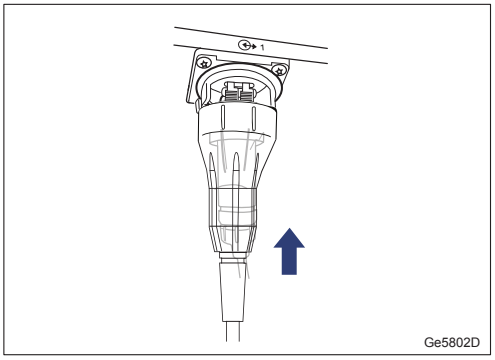
Ensure the optical cable connectors and SFP modules are clean. Always clean the optical connectors with the cleaning cassette before installing the cables in the SFP module. Only clean the SFP module if needed.

6. 

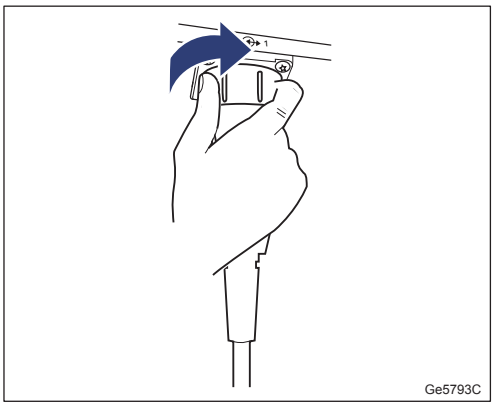
Plug in the connector into the SFP receptacle in the optical connection port.

7. 

Clamp the plug body with the seal ring to the optical cable.

8. 

Slide the bayonet shell to the bulkhead connector.

9. 

Clamp the bayonet shell on the bulkhead connector clockwise.

After This Task

When the installation of the AIR is finished, strap the cables to the pole or wall.

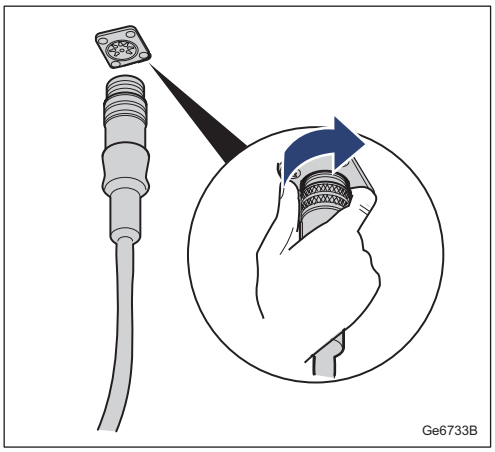
5.4 Connect EC Light Cable (Optional)

How to connect the EC light cable to the unit.

Prerequisites

If using more than one DIN 14-pin function, a Y-cable must be connected to the DIN 14 connector for each added function.

Steps

1. Remove the protective cap from the DIN 14-pin interface.
2. 

Connect the EC light cable.

After This Task

When the installation of the AIR is finished, strap the cables to the pole or wall.

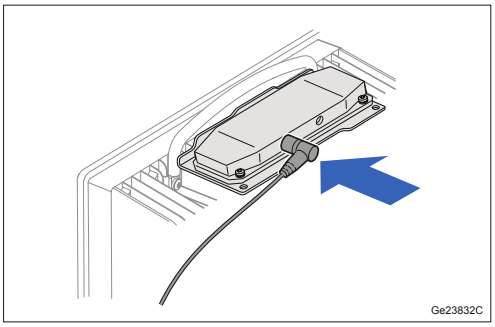
5.5 Connect RAE Cable (Optional)

How to connect the RAE cable.

Prerequisites

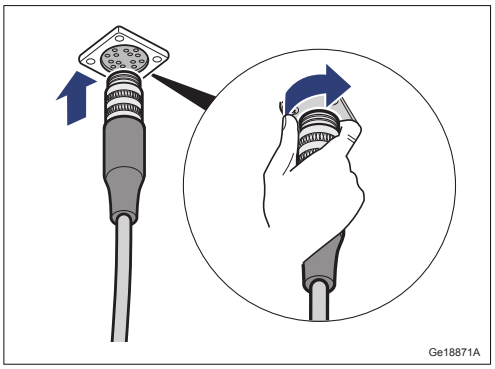
If using more than one DIN 14-pin function, a Y-cable must be connected to the DIN 14 connector for each added function.

Steps

1. 

   Connect the angled connector on the RAE cable to the RAE unit.

2. Remove the protective cap from the DIN 14-pin interface.

3. 

   Connect the RAE cable.

After This Task

Route the cable on the left side of the first cooling fin on the unit. Fasten the cable on top and bottom of the unit with the delivered straps.

5.6 Connect External Alarm Cable (Optional)

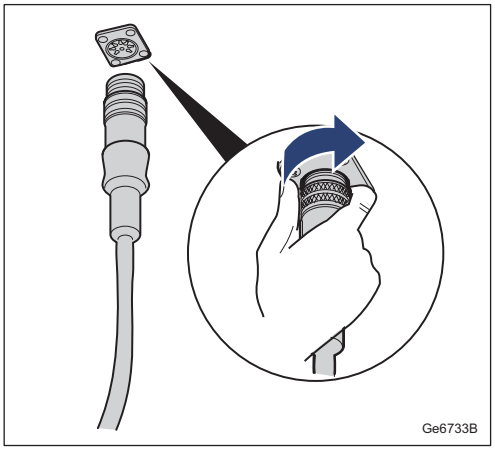
> How to connect the external alarm cable to the unit.

Prerequisites

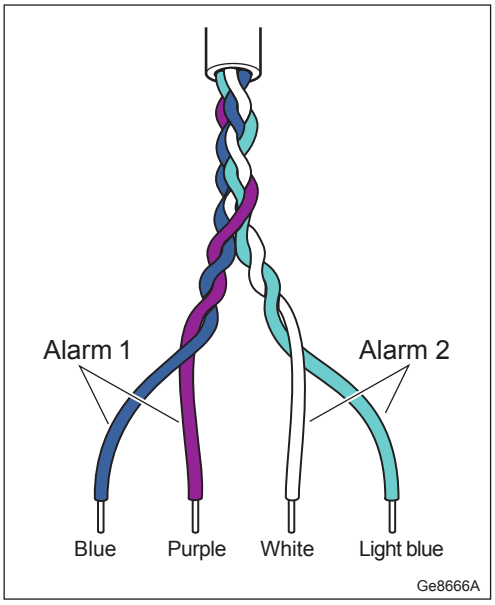
If using more than one DIN 14-pin function, a Y-cable must be connected to the DIN 14 connector for each added function.

Steps

1. Remove the protective cap from the DIN 14-pin interface.

2. 

Connect the external alarm cable to the unit.

3. 

Connect the other end of the external alarm cable to the external equipment, using the following cable data:

5.7 Connect TX Monitor Cable (Optional)

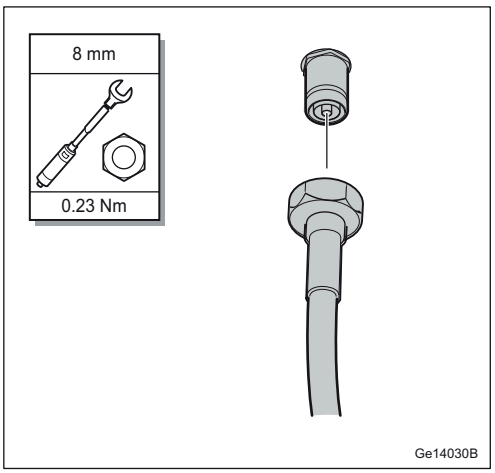
How to connect the TX monitor cable to the unit.

Required Tools and Equipment

Tool	Product Name
	Torque wrench — 0.23 Nm
	Open-ended wrench head

Tool	Product Name
	— 8 mm

Steps

1. Remove the shielding cap from the TX monitor interface.
2. 

Connect the TX monitor cable.

After This Task

When the installation of the AIR is finished, strap the cables to the pole or wall.

5.8 Connect Power Cable

> How to connect the power cable.

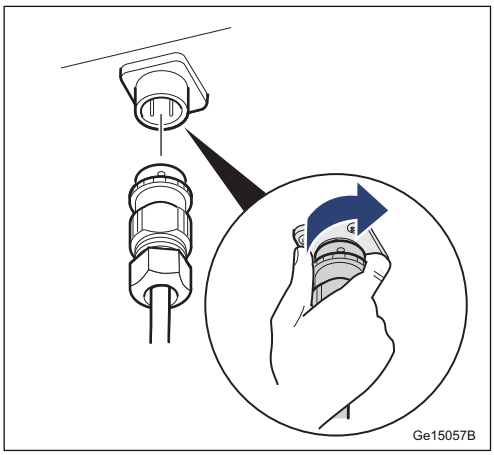
Prerequisites

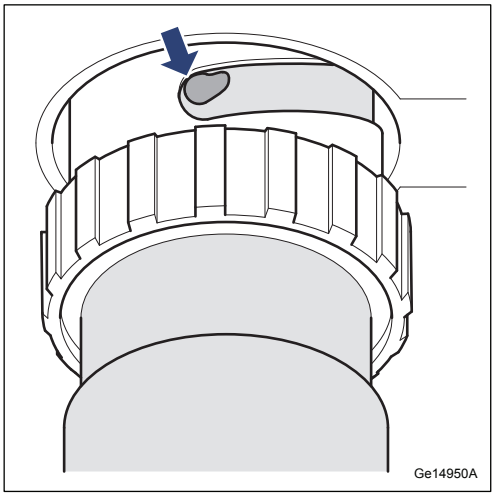
Warning!

High energy levels are present in this unit. Improper handling of the unit can lead to short circuits that can result in serious injury. Exercise care when working with this unit.

Steps

Note: Make sure that all connectors are in good condition and replace if needed. Model number is printed on the power connector. For 3-wire power connector use model 7F2RS103B-0M16T1(RNT 447 36/01). For 2-wire power connector use model 7F2RS104B-0M16T1(RNT 447 37/01).

1. Remove the protective cap from the power interface.
2. 

   Connect the power cable for 3-wire (DC-I) or 2-wire (DC-C) connections.
3. 

   Make sure that the connector is locked correctly to prevent bad connection and impact from rain or humidity. The figure shows the correct lock position.

After This Task

When the installation of the AIR is finished, strap the cables to the pole or wall.

6 Verify Installation

> After the installation is completed, the grounding system, the AIR, and the power supply system need to be verified.

Check Grounding System

Table 1 Grounding Checklist

#	Item	Yes	N/A (1)
1	Grounding system installed as specified in *Customer Site Documentation*?		
2	The unit grounding installed and connected to site MET?		
3	The unit grounded, washers in place, and bolts tightened to a torque of 8 Nm?		
4	Grounding lugs, terminals, C-clamps, and cable lugs correctly crimped?		
5	Ground cable strapped at least once every meter?		
6	Ground cable insulation undamaged?		
7	Ground cable cross-section is at least 16 mm^2 for copper cable, or a similar cable?		
8	Equipment is undamaged and free from contamination?		
9	Protective grounding in power supply box connected to site MET?		
Notes:			

(1) Not Applicable

Check the AIR

Table 2 AIR Checklist

#	Item	Yes	N/A
1	AIR installed as specified in *Customer Site Documentation*?		
2	AIR pole clamps compatible with the pole (no risk for corrosion)?		
3	The AIR properly aligned horizontally and vertically?		

#	Item	Yes	N/A
4	The AIR is clean and not damaged?		
5	The AIR side cooling vents are free of debris?		
6	No damaged cables or connectors?		
7	Connectors properly connected?		
8	Optical cables properly connected, labeled, and strapped? (1)		
9	Optical cables lead-in sealed?		
10	Bend or drip loop applied on optical cables outside wall or roof?		
11	Protective covers, dust caps, and terminations plugs are installed on unused ports?		
12	Outdoor cables ladder protected from falling ice?		
13	Equipment labeled as required as specified in *Site Installation Documentation*?		
Notes:			

(1) Ensure that the minimum bending radius has not been violated. For more information, see Handling SFP Modules and Optical Cables.

Check Power Supply System

Table 3 Power Supply System Checklist

#	Item	Yes	N/A
1	AIR installed as specified in the site installation documentation?		
2	AIR labeled correctly and terminated in the right position?		
3	Power socket is securely and correctly attached to the AIR?		
Notes:			

7 Power Up AIR

> After the installation has been verified, the unit can be powered up and the optical indicators will show the status of the unit.

Prerequisites

The AIR must be powered up within 48 hours after installation even if the AIR is not taken into service. The reason is to protect the internal units from local climate conditions.

Do!

Products not connected to ground risk damage from overvoltage or overcurrent. Always ground products according to instructions.

Caution!

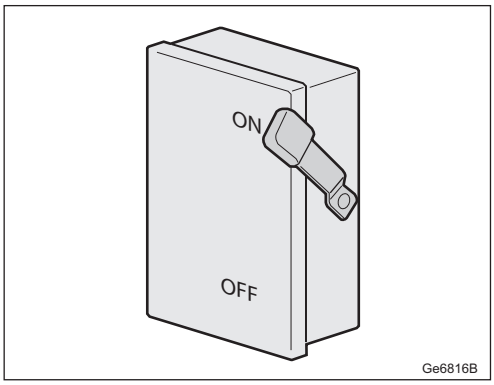
Do not stand or work in front of an operational antenna, unless it has been verified or documented that RF exposure levels are within specified safety limits.

Caution!

Excessive RF exposure can result in potentially adverse health effects. If it is suspected that RF exposure limits may be exceeded, ensure that transmitting antennas are switched off, or reduce output power whilst working with or near antennas.

For more information about RF EMF exposure, see Antenna Integrated Radio Unit Description and Radio Frequency Electromagnetic Fields.

Steps

1. 

   Switch on the external power to the AIR.

Results

Once the power is switched on and the AIR is connected to baseband, the AIR is loaded with the AIR-specific software. Optical indicators show the status of the AIR.

Note: It can take up to 10 minutes or more before the AIR is in operation.

After This Task

Check the status of the optical indicators, see Table 4.

Table 4 AIR Optical Indicators

Marking	Indicator	Color	Correct Status
!	Fault	Red	Off
✔	Operational	Green	Flashing or On
🔧	Maintenance	Blue	Off
⊙→	Optical	Green	On

For detailed information about optical indicators, see Antenna Integrated Radio Unit Description and Indicators, Buttons, and Switches.

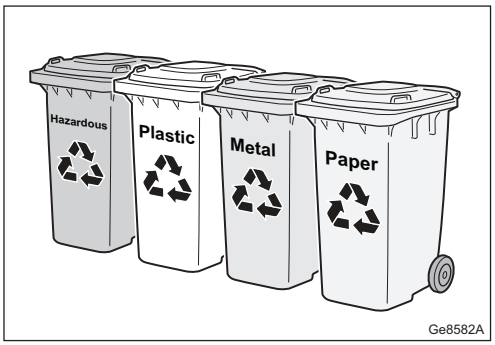
Note: After the radio power is switched on, it takes 20 to 30 seconds before the slow flashing (0.5 Hz) starts. The operational indicator only flashes once during this period.

The sliding cover can be closed to hide the optical indicators. Depending on the work order, leave the sliding cover open or closed.

After the power-up, it is recommended to run the installation self-check, as described in Perform Basic Stand-alone Radio Installation Check. This feature helps to identify potential connection faults or hardware faults of the AIR.

8 Perform Concluding Routines

After successful installation, the site needs to be cleaned and the customer site documentation needs to be handled.

1. Clean the site and remove objects such as wrapping paper and cable clippings.
2. 

Dispose of waste according to local regulations.
3. Fill in the verification checklist in the *Customer Site Documentation*.
4. Report any faults according to local requirements.
5. Hand over the *Customer Site Documentation* to the person responsible for the site.
6. Lock all doors and gates to the site.

8.1 Environment

Ericsson strongly recommends that installers pay particular attention to the environment when cleaning the site after installation.

Recycle all waste that can be recycled and sort the rest so that they can be disposed of according to local regulations. Use the checklist in Table 5 for recycling and sorting.

Table 5 Recycling and Sorting Waste

Recycle or Sort	Item	YES	N/A (1)
Sorted with plastics	Pieces of cable with low metallic content		
	Cable insulation from crimping, brazing, or welding		

Recycle or Sort	Item	YES	N/A (1)
	Polyethylene		
	Bubble plastic		
	Cable tie clippings		
Paper recycling	Paper		
Recycled or sorted as wood	Wood		
Recycled or sorted as metals	Waste metal from cable ladders		
	Pieces of cable with high metallic content		
	Nuts, bolts, washers, and screws		
Notes:			

(1) Not Applicable

applied sciences

MDPI

Review

Probing Ferroic States in Oxide Thin Films Using Optical Second Harmonic Generation

Johanna Nordlander, Gabriele De Luca, Nives Strkalj, Manfred Fiebig and Morgan Trassin *

Department of Materials, ETH Zurich, Vladimir-Prelog-Weg 4, 8093 Zurich, Switzerland; johanna.nordlander@mat.ethz.ch (J.N.); gabriele.deluca@mat.ethz.ch (G.D.L.); nives.bonacic@mat.ethz.ch (N.S.); manfred.fiebig@mat.ethz.ch (M.F.)

* Correspondence: morgan.trassin@mat.ethz.ch

Received: 28 February 2018; Accepted: 3 April 2018; Published: 6 April 2018

Abstract: Forthcoming low-energy consumption oxide electronics rely on the deterministic control of ferroelectric and multiferroic domain states at the nanoscale. In this review, we address the recent progress in the field of investigation of ferroic order in thin films and heterostructures, with a focus on non-invasive optical second harmonic generation (SHG). For more than 50 years, SHG has served as an established technique for probing ferroic order in bulk materials. Here, we will survey the specific new aspects introduced to SHG investigation of ferroelectrics and multiferroics by working with thin film structures. We show how SHG can probe complex ferroic domain patterns non-invasively and even if the lateral domain size is below the optical resolution limit or buried beneath an otherwise impenetrable cap layer. We emphasize the potential of SHG to distinguish contributions from individual (multi-) ferroic films or interfaces buried in a device or multilayer architecture. Special attention is given to monitoring switching events in buried ferroic domain- and domain-wall distributions by SHG, thus opening new avenues towards the determination of the domain dynamics. Another aspect studied by SHG is the role of strain. We will finally show that by integrating SHG into the ongoing thin film deposition process, we can monitor the emergence of ferroic order and properties in situ, while they emerge during growth. Our review closes with an outlook, emphasizing the present underrepresentation of ferroic switching dynamics in the study of ferroic oxide heterostructures.

Keywords: multiferroic; ferroelectric; SHG; domain engineering; thin films; $BiFeO_3$; PZT; domain walls

1. Introduction

Among the ferroic systems, ferroelectrics are subject to intense scrutiny by the scientific community due to their promising impact on current and future energy-efficient technologies [1–4]. The low-energy consuming electric field driving their polarization direction, and therefore their functionality, makes ferroelectrics of interest for novel device applications. In the drive for power-efficient oxide electronics, the recent development regarding the control of thin film ferroelectricity spans from the determination of critical thicknesses in the ultrathin regime [5,6] to ferroelectric domain engineering using epitaxial strain [7,8]. Since all ferroelectrics are piezoelectric, their range of applications can even be extended beyond charge-based memory devices to include mechanical device concepts such as stress sensors, etc. [2]. Furthermore, in the case of coexisting electric and magnetic orders in multiferroic thin film materials, new avenues toward increased data storage density can be realized [9–11]. A memory bit could indeed be stored independently in the magnetic and in the ferroelectric state. In addition, when the two ferroic orders are coupled, the

magnetoelectric coupling can enable the implementation of energy-efficient, electrically writable magnetic memories [12–15].

Reviews on SHG and its potential for probing ferroic states in bulk crystals are available in the literature [16–18]. Here we place the emphasis on recent development in SHG for investigating ferroelectric and multiferroic thin films and heterostructures. Therefore, the first part of this review is devoted to ferroic domain design in thin layers and superlattices. In the following sections, we review the progress made in SHG as tool to access the properties and challenges posed by ferroelectric and multiferroic domains in thin film architectures. By its symmetry sensitivity and its non-invasive nature, SHG appears as a complementary tool to the traditional approach by scanning probe microscopy (SPM) and electron microscopy.

1.1. Ferroelectric Thin Films

While tremendous advances have been made in understanding ferroelectrics on their own, fundamental aspects of their behavior in the ultrathin regime or once inserted in multilayers or superlattice architectures are still under intense investigation [19–21]. The buildup of a ferroelectric polarization P in a material is accompanied by the accumulation of bound charges at its surfaces. As a result, the system hosts an internal field, induced by these bound charges, which is oppositely oriented to the polarization direction; see Figure 1. This depolarizing field D depends on the thickness of the ferroelectric layer and its electrostatic environment, such as the presence of metallic electrodes or the surrounding gas partial pressure and their respective charge screening efficiencies [22–24]. The role of the depolarizing field becomes critical in ultrathin ferroelectric single layers and in superlattices, where interface effects may become predominant [19,21,25,26]. As illustrated in Figure 1, a small change in film thickness and charge environment will have a drastic influence on the manifestation of a ferroelectric state and its domain configuration. For instance, in superlattices, polarization can be stabilized at ultra-low thickness [27,28] or a global enhancement of ferroelectricity can be achieved [29]. In a charge compensating environment, a single domain state is favored in a wide range of thicknesses; see Figure 1b. In the case of absent or incomplete charge screening, the presence of the depolarizing field will result in splitting of the single-domain state into a 180° multi-domain pattern, the formation of flux-closure domains or the emergence of ferroelectric vortices depending on the thickness and strain state considered; see Figure 1c.

In ferroelectric-based heterostructures, the intricate character of the interactions between polar and non-polar, metallic and non-metallic layers and the difficulty to access buried ferroelectric states renders the control of polarization challenging. Furthermore, probing the emergence of ferroelectricity or the domain state in thin films as function of thickness, strain and temperature is hindered by the lack of contactless probing techniques. Leakage currents [30] in the ultrathin regime and at elevated temperatures prevent the understanding of ferroelectricity in the low thickness regime.

On the structural level, pioneering theory and experimental work dealing with ferroelectric thin film behavior under epitaxial stress evidenced the potential of epitaxial strain as a new degree of freedom for controlling the ferroelectric state in single layers [7]. In situ oxide growth diagnostic tools such as reflection high energy electron diffraction (RHEED) [31] together with the development of single crystal substrate manufacture, allow precise engineering of the strain state in high-quality thin films. Using a combination of phase field simulations and experimental techniques, it was further shown that strain can be used to tune the ferroelectric polarization [32], enhance the Curie temperature [33,34] and induce a selectivity to specific ferroelectric switching events [35]. Epitaxial strain can also be used to induce a ferroic state. For instance, ferromagnetism was strain induced in orthorhombic $LuMnO_3$ thin films [36] while a polar state was stabilized in tensile strained $SrMnO_3$ thin films [37]. However, experimental evidence of such a control remains inaccessible in multilayer architectures due to the increasing role of interfaces and the lack of a direct probe of ferroelectricity in a buried configuration.

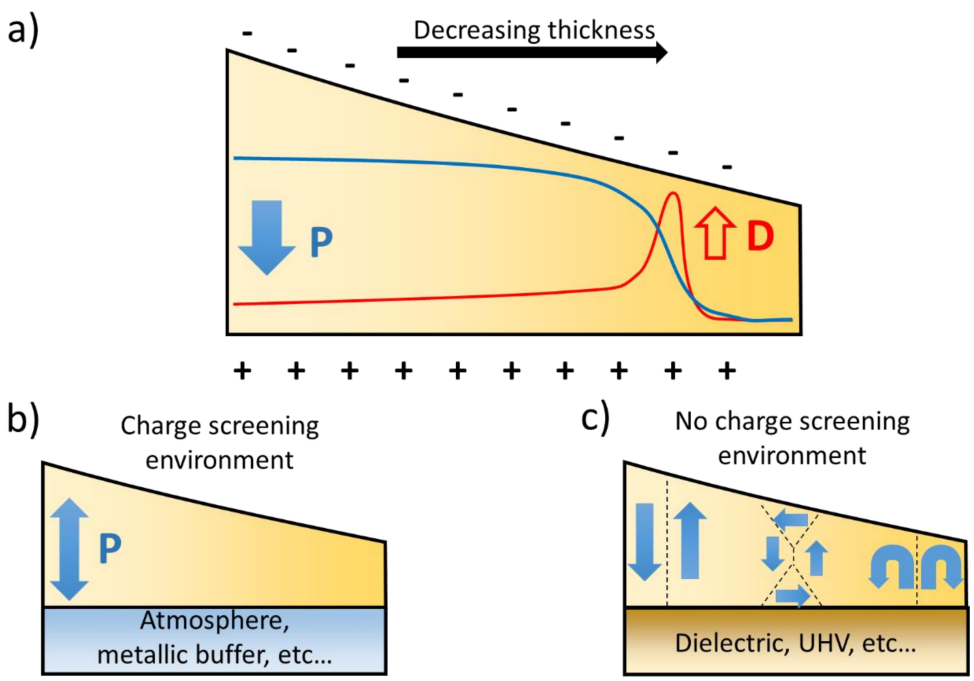


Figure 1. Thickness dependence of the polar state in ferroelectric thin films. (**a**) The polarization *P* and depolarizing field *D* have a thickness dependence, resulting in the minor influence of the depolarizing field on polarization in the high thickness range and a suppression of polarization at low thicknesses; (**b**) When the bound charges can be screened by the environment, the depolarizing field is minimized and a single domain state can be stabilized over a wide thickness range; (**c**) In the absence of charge screening, however, the depolarizing field favors a multi-domain state. Flux closure or vortices can also form in a specific thickness or strain regime.

1.2. Multiferroicity in Thin Films

The so called multiferroic materials are defined in the single-phase case by the simultaneous presence of two or more ferroic orders. Hence, a ferroelectric state can under certain conditions coexist with magnetic order [38]. Multiferroic materials were first investigated in the late 1950s and were the subject of an intense revival when the scientific community became aware of their high potential towards low-energy consuming electronics [39,40].

Taking advantage of the progress in thin film deposition techniques, the early motivation for multiferroic thin-film growth resided in reproducing multiferroic bulk properties in reduced dimensions to lower the external switching field requirements for applications. The pioneering results in the field deal with the growth of multiferroic magnetoelectric $BiFeO_3$ (BFO) thin films [41] and multiferroic hexagonal manganite thin films [42,43]. In both cases, the coexistence of magnetic ordering and ferroelectric polarization was shown. The multiferroic thin film research field subsequently evolved and led to the development of strain engineering of multiferroic properties [44–46] and the design of artificial multiferroic heterostructures where multiferroicity is introduced as a product effect by the combination of different ferroic constituents [44].

Following the recent development of the electrical control of magnetization using multiferroics [47], the development of experimental techniques giving access to multiferroic and magnetoelectric domain configurations has become a priority for the scientific community.

1.3. Ferroic Domain Engineering in Single Layers and Superlattices

The combination of epitaxial strain, surface termination and electrostatics in ferroic thin films and heterostructures can be used for domain engineering. Deterministic control of the domain formation

process toward the design of complex ferroic states in three dimensions is a necessary means to realize next-generation oxide-based devices [48–51].

One instructive example of strain engineering the ferroic domain structure in single layers is the case of multiferroic BFO thin films. In this system, Chu and coworkers [52] demonstrated the ability to modify the ferroelectric domain state by using relevant substrate orientations and surface states. Starting from the eight symmetry-allowed polar axes along the <111> crystallographic directions of the pseudo-cubic unit cell, the number of as-grown ferroelectric variants could be tuned to either a four ferroelectric variants or a two ferroelectric variants configuration, when grown on the cubic perovskite (001) or (110) orientation of $SrTiO_3$ (STO) substrates, respectively. A single domain state could be achieved when using (111) oriented STO (see Figure 2a–c). Further tuning of the domain architecture was achieved with the use of scandate substrates [53]. $DyScO_3$ (DSO) crystallizes in the orthorhombic space group with a pseudo-cubic unit cell exhibiting a slight anisotropy between the in-plane lattice parameters [54]. This results in an in-plane anisotropic strain which can further lift the degeneracy of polar domains of perovskite ferroelectrics [53]. Scandate substrates have also been used to induce multiferroicity in STO thin films [55,56] and $EuTiO_3$ layers [57,58].

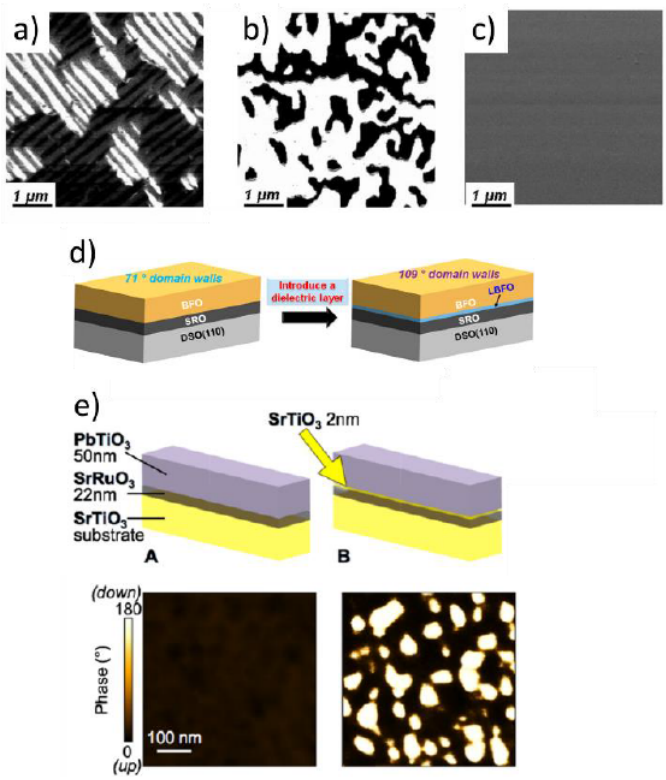


Figure 2. Domain engineering in ferroic thin films. (**a**–**c**) Control of the number of ferroelectric variants in BFO using STO substrate orientation. (001), (110) and (111) orientations correspond to (**a**–**c**), respectively. Reproduced with permission from [52], Copyright Wiley-VCH Verlag GmbH and Co. KGaA, 2007); (**d**) The insertion of a dielectric ultrathin lanthanum-doped BFO layer between BFO and the metallic $SrRuO_3$ (SRO) buffer induces a transition from a 71° stripe domain architecture to a 109° stripe domain pattern. Reprinted with permission from [62], Copyright American Chemical Society, 2016; (**e**) The insertion of a dielectric STO layer between the SRO and $PbTiO_3$ (PTO) ferroelectric film stabilizes nanoscale 180° domains. Reprinted with permission from [64], Copyright American Chemical Society, 2014.

Beyond epitaxial strain considerations alone, the macroscopic out-of-plane polarization orientation and the domain state in thin film ferroelectrics also drastically depend on their electrostatic

environment [59] and the surface termination of the substrate [60,61]. Screening conditions affect the stability of domain walls and are a key parameter in the control of complex domain structures. By inserting a dielectric layer between the ferroelectric thin film and the metallic buffer, the 109° stripe domain state could be favored in BFO [62] and 180° domains were stabilized in $PbTiO_3$ (PTO) [63,64]; see Figure 2d,e, respectively.

By tuning ferroelectric layer thicknesses and controlling the nature of the charge screening environment in heterostructures and superlattices, it is possible to engineer the ferroelectric domain architecture in three dimensions [61,65]. A prominent example is the formation of ferroelectric vortex domains in PTO/STO superlattices, as shown in Figure 3a,b [48] or the stabilization of flux-closure domain patterns in PTO single layers inserted between two STO dielectric layers; see Figure 3c [65].

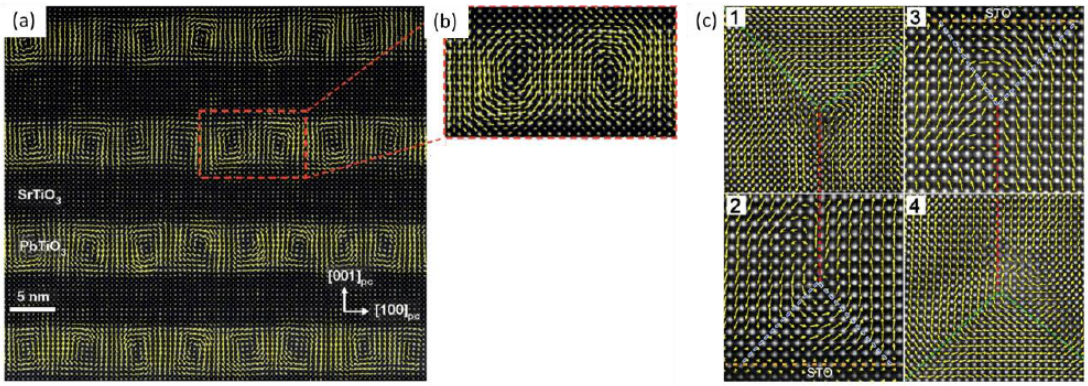


Figure 3. Scanning transmission electron microscopy on complex polar domains. (**a**) Observation of ferroelectric vortices in PTO/STO superlattice. Reprinted by permission from [48], Copyright Macmillan Publishers Ltd., 2016. (**b**) Close-up into a single vortex. (**c**) Flux-closure domain patterns in a PTO/STO multilayer system. From. Reprinted with permission from [65], Copyright AAAS, 2015.

The direct observation of three-dimensional ferroelectric and multiferroic domain architectures is however not trivial. The conventional SPM technique usually requires the contact with the functional layer and is inherently surface sensitive, therefore lacking the demanded depth analysis of the structure. A particular challenge arises in probing domain structures in a multilayer configuration where the ferroic layers are capped by an impenetrable layer. In active devices, the functional layer is often capped by a conducting electrode which limits the piezoresponse force analysis to the out-of-plane polarization component [4,66,67].

1.4. Current Challenges in Probing Ferroic Domains in Thin Films

Currently, state-of-the-art observation of ferroelectric domain distribution in the volume is provided primarily by scanning transmission electron microscopy (STEM). This imaging technique allows spatial mapping of local polarization but remains a destructive analysis and is highly dependent on the sample preparation. Other experimental techniques such as synchrotron X-ray diffraction-based techniques [21,26,68] can identify polarization orientation and additional periodicities originating from domain formation, but the involvement of large-scale facilities renders their generalization towards day-to-day operation impractical. For a complete understanding of the role of epitaxial strain and electrostatics at the created oxide interface, noninvasive access to the polarization state is essential.

When it comes to probing several ferroic orders simultaneously, as is desirable in multiferroics, additional experimental constraints need to be considered. Most ferroelectric multiferroic materials exhibit antiferromagnetic ordering and therefore do not possess a net magnetic moment. Hence, antiferromagnetic domain visualization remains challenging. An efficient way to observe antiferromagnetic domains is using the principle of ferroic domain imprint transfer [69–73]. A thin ferromagnetic layer is deposited in direct contact with the multiferroic layer. The interfacial exchange

coupling with the underlying antiferromagnetic layer results in the formation of ferromagnetic domains in the ferromagnetic film that can be resolved using scanning electron microscopy with polarization analysis (SEMPA) [72] as shown in Figure 4a,b, or magnetic force microscopy (MFM) [72,74]. The underlying ferroelectric domain state is however no longer reachable using conventional scanning probe experiments, i.e., piezoresponse force microscopy (PFM), after the ferromagnet has been deposited.

Recently, a single spin magnetometer technique was developed based on point-like impurity nitrogen-vacancy (NV) defect in diamond [75–77]. This scanning-probe technique was used to spatially resolve the local magnetic order in multiferroic antiferromagnetic BFO thin films [78]. A periodic modulation in the magnetic response corresponding to a spin cycloid was measured within a single ferroelectric domain; see Figure 4c,d. This spatially resolved characterization of antiferromagnetic domains does not prevent the subsequent ferroelectric analysis in the same surface area and signifies important progress towards the analysis of multiferroic domains.

However, the direct access to multiferroic domains in buried configurations, i.e., in a device architecture or in a superlattice, remains impossible using the scanning probe approach.

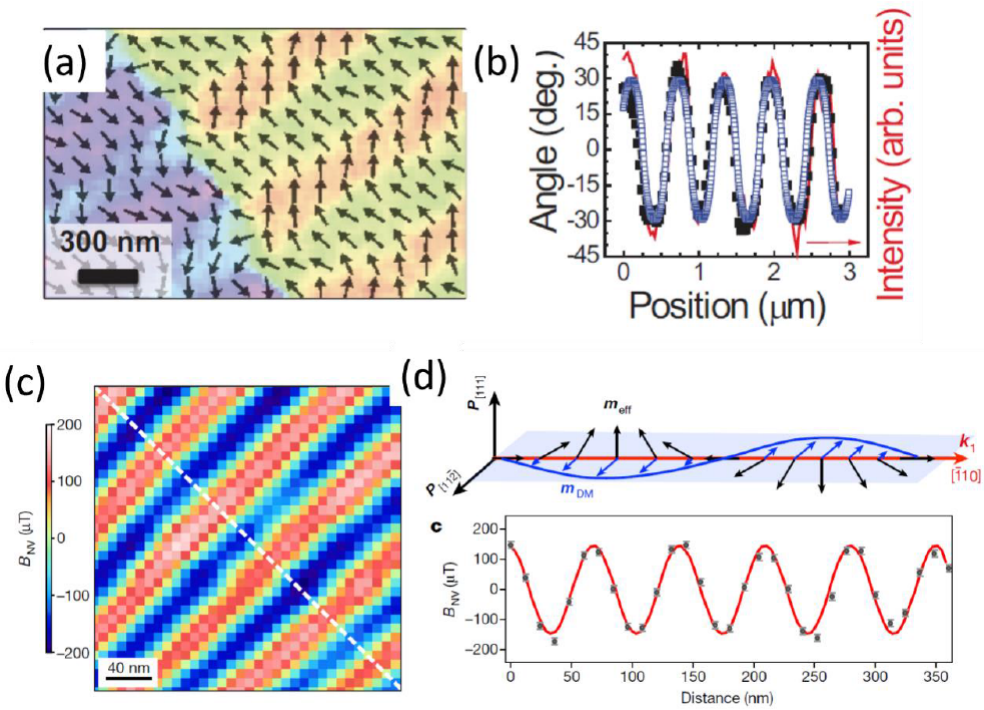


Figure 4. Imaging multiferroic domains in BFO thin films. (**a**) SEMPA showing the in-plane orientation of the ferromagnetic domains in a CoFe layer deposited on top of the multiferroic film; (**b**) Line scans from the backscattering electrons (red curve), SEMPA (filled black squares) and calculations (open blue squares) with a 7 mT effective coupling field highlight the coupling between the multiferroic and the ferromagnet. (**a**,**b**) Reprinted with permission from [72], Copyright the American Physical Society, 2013; (**c**) Magnetic field distribution recorded on a single ferroelectric domain showing the existence of a spin cycloid within a single ferroelectric domain; (**d**) Representation of the spin density wave (top) showing an uncompensated moment, the line scan (bottom) shows the magnetic field distribution along the cycloid propagation direction. Reprinted by permission from [78], Copyright Macmillan Publishers Ltd.: Nature, 2017.

2. Optical Second Harmonic Generation for Investigation of Ferroic Materials

Optical second harmonic generation (SHG) offers a complementary approach to SPM analysis in ferroic analysis. It is non-invasive, contact-free and can probe the entire volume of a thin film structure including buried layers in heterostructures.

2.1. SHG in Bulk Ferroic Systems

Optical SHG refers to the frequency doubling of a light wave. The SHG emission in a crystal depends on its point group symmetry and SHG is therefore in general sensitive to any point symmetry violation in the material. Since the emergence of ferroic order is intimately linked to a lowering of the crystal point-group symmetry, such ordered states are aptly probed with SHG [16–18,79]. More importantly, this technique can be used to probe *coexisting* ferroic states, as in multiferroics. The source term for SHG is described macroscopically by the tensor Equation (1).

$$P_i(2\omega) = \varepsilon_0 \chi_{ijk} E_j(\omega) E_k(\omega) \tag{1}$$

The tensor components χ_{ijk} belong to the third-rank second-order susceptibility tensor $\chi^{(2)}$, parameterizing the non-linear light-matter interaction. The indices i, j and k are defined with respect to the crystallographic axes of the crystal x, y and z and denote the directions of the respective oscillating fields. E is the electric field component of the incident light at frequency ω, and P is the polarization at frequency 2ω induced inside the material. This induced polarization acts as a source of an emitted, frequency-doubled, light wave with intensity $I_{SHG} \propto |P\ (2\omega)|^2$. The point-group symmetry of a compound determines its set of tensor components $\chi_{ijk} \neq 0$. The allowed tensor components for each symmetry group can be found tabulated in literature [80].

Experimentally, specific χ_{ijk} components can be accessed through a careful choice of incident and detected light polarization [16–18]. The spectral dependence of the $\chi^{(2)}$ tensor can be determined using a tunable wavelength light source and can provide in-depth information on specific electronic transitions and optimal SHG response [81,82]. A schematic of a basic SHG set up in transmission is given in Figure 5. The direction of polarization of the probe beam is defined by the polarizer angle, and the polarization direction of the detected SHG light is projected along the analyzer angle, as indicated in Figure 5. Because the second harmonic light is spectrally distinct from the fundamental light, the SHG frequency can be simply monitored separately from the probe beam intensity and is as such a background-free characterization technique.

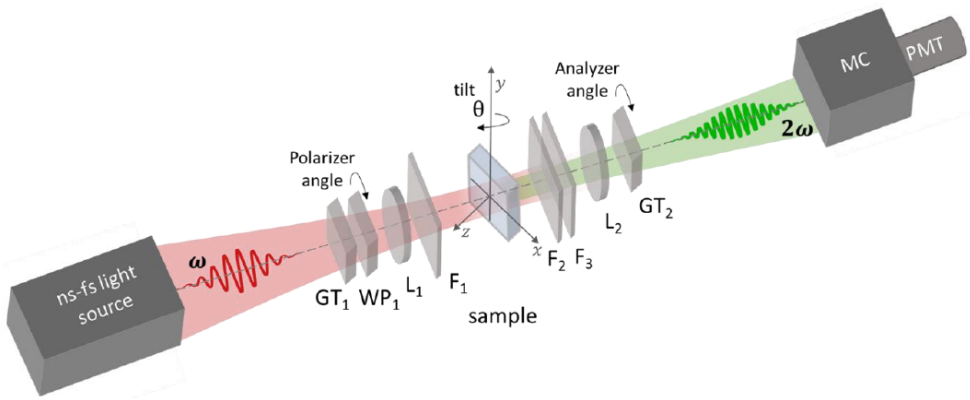


Figure 5. Typical SHG measurement setup in transmission. The incident probe beam is linearly polarized by a Glan-Taylor prism (GT_1). The polarization of the probe beam is then set to an arbitrary polarization state by a rotatable wave-plate. WP_1 can be a half-wave-plate for linear polarization states or a quarter-wave-plate to achieve circular polarization. The spot size on the sample is controlled by a focusing lens L_1. Low-pass filter F_1 removes any second harmonic light generated in the polarization optics before the sample. The fundamental beam and higher order harmonics are blocked by band/high-pass F_2 and low-pass F_3 filters. The SHG light is collected by L_2, and the polarization state is analyzed by projection on a second, rotatable, GT prism (GT_2). The SHG intensity is integrated after passing a monochromator (MC) and photomultiplier tube (PMT). In the case of spatially-resolved SHG, L_2 is exchanged for a (microscope) objective lens and the detection unit is replaced with a CCD camera. The incident angle θ can be changed by rotating the sample around the y-axis.

In the specific case of ferroelectrics, inversion symmetry is broken by a polar distortion, allowing non-zero χ_{ijk} components. To characterize the polar distortion along a certain crystallographic direction, and to be able to discriminate between excited tensor components, one can fix the detected polarization direction of the second harmonic light, i.e., the analyzer angle, while rotating the polarization of the probe beam. Here, we refer to such characterization as polarizer measurements. In general, the polar axis of the investigated sample can point along an unknown direction. This orientation can be accessed by fixing both the incoming and detected light polarization while rotating the sample along the azimuthal angle defined by the sample surface.

Equivalently, this in-plane polar anisotropy measurement can also be performed with a simultaneous rotation of the polarization of both probing and detected light while keeping the sample position fixed. Because the SHG technique only probes polar distortions projected on a plane perpendicular to the k-vector of the incident light wave, it may be necessary to introduce a tilt of the angle of incidence (denoted by θ in Figure 5). This changes the direction of the k-vector with respect to the crystal and provides access to certain domain orientations.

SHG is a well-established technique to non-invasively probe bulk materials where domain dimensions are expected to be larger than the spatial resolution limit (about a micrometer). Indeed, it has been used to image ferroelectric domains [83], antiferromagnetic domains [84,85] and more recently multiferroic domains in bulk systems [86,87], as can be seen in Figure 6.

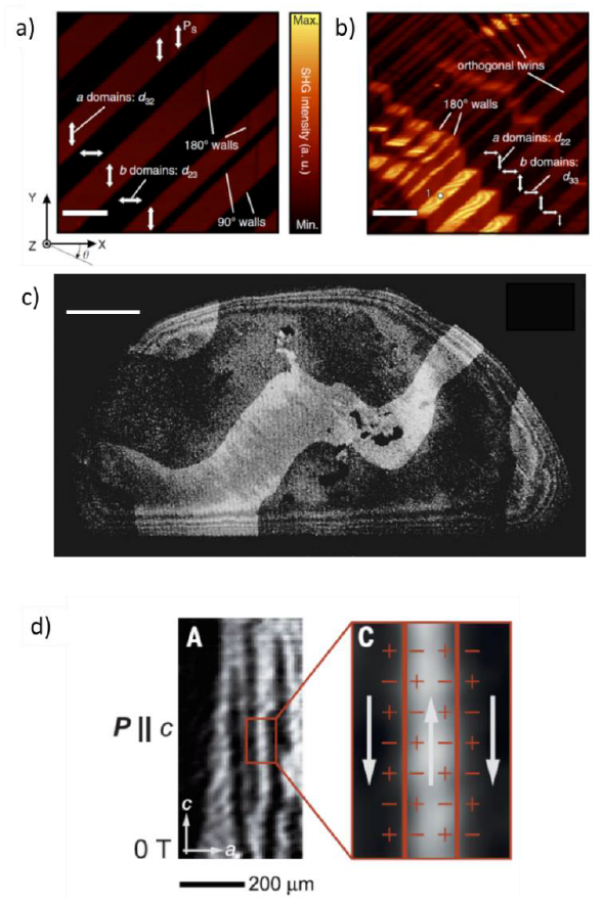


Figure 6. Imaging ferroic domains in bulk using SHG. (**a**) Ferroelectric domain structure in $BaTiO_3$ (BTO). The scale bar is 8 µm. (**b**) Observation of orthorhombic twinning in a BTO crystal. The scale bar is 12 µm. Reprinted by permission from [83], Copyright Macmillan Publishers Ltd.: Nature Communications, 2014. (**c**) Antiferromagnetic domains in Cr_2O_3, reprinted with permission from [84] Copyright AIP Publishing, 1995. Scale bar is 1 mm. (**d**) Multiferroic domains in $TbMnO_3$, Reprinted with permission from [86], Copyright AAAS, 2015.

Local symmetry mapping is performed by selectively coupling to SHG tensor components which are domain orientation dependent. Spatially-resolved SHG, using a charge-coupled device (CCD) camera detection system, then shows domains with 90° relative orientation in different brightness; see Figure 6a. Since each phase of a material corresponds to an individual set of non-linear tensor components, phase coexistence can also be mapped using optical SHG. Figure 6b demonstrates the presence of orthorhombic twins in a tetragonal $BaTiO_3$ (BTO) crystal. Furthermore, domains with opposite order parameter, i.e., $\pm P$ in the ferroelectric case, emit SHG light waves with a relative 180° phase shift. Superimposing these light waves with a reference light, one can achieve domain contrast and observe oppositely oriented domains with different brightness [79,88].

Moreover, the emergence of SHG signal as a function of temperature is a powerful method to identify the critical temperature of phase transitions in the ferroic compound under investigation. This is particularly useful when standard characterization techniques are unable to characterize the emergence of the ferroic order [86].

For exemplifying the relation that connects a particular pattern of ferroic domains to a specific polarization-dependent SHG signal we consider the case of a single-domain ferroelectric state in the multiferroic model system BFO. The point symmetry of ferroelectric BFO allows a set of independent non-zero SHG susceptibilities. These are described as complex values so that they possess an amplitude χ_{ijk} and a relative phase Ψ_{ijk}. The resulting ferroelectric SHG signal is then given by [88]:

$$|P_i(2\omega)|^2 \propto \left|\sum_{j,k} \chi_{ijk} e^{i\Psi_{ijk}} E_j(2\omega) E_k(2\omega)\right|^2 \tag{2}$$

With z as surface normal, we find χ_{xxx}, χ_{xxy}, χ_{xyy}, χ_{yyy}, χ_{yyx}, χ_{yxx}, χ_{xzz}, χ_{yzz}, χ_{zzx}, χ_{zzy} as non-zero independent components for BFO (and accordingly for Ψ_{ijk}) in accordance with its point symmetry group m. The effective polarization components of the SHG wave are given by

$$P_x(2\omega) \propto \chi_{xxx} e^{i\Psi_{xxx}} \cos^2\theta \sin^2\varphi + 2\chi_{xxy} e^{i\Psi_{xxy}} \cos\theta \cos\varphi \sin\varphi + \chi_{xyy} e^{i\Psi_{xyy}} \cos^2\varphi + \chi_{xzz} e^{i\Psi_{xzz}} \sin^2\theta \sin^2\varphi \tag{3}$$

$$P_y(2\omega) \propto \chi_{yyy} e^{i\Psi_{yyy}} \cos^2\varphi + 2\chi_{yyx} e^{i\Psi_{yyx}} \cos\theta \cos\varphi \sin\varphi + \chi_{yxx} e^{i\Psi_{yxx}} \cos^2\theta \sin^2\varphi + \chi_{yzz} e^{i\Psi_{yzz}} \sin^2\theta \sin^2\varphi \tag{4}$$

$$P_z(2\omega) \propto 2\chi_{zzx} e^{i\Psi_{zzx}} \cos\theta \sin\theta \sin^2\varphi + 2\chi_{zzy} e^{i\Psi_{zzy}} \sin\theta \cos\varphi \sin\varphi \tag{5}$$

Here, θ is the tilting angle, defined as the angle between the surface normal of the film and the k-vector of the incident light ($\theta = 0°$ corresponds to normal incidence). Furthermore, φ is the angle between the polarization of the linearly polarized incoming light and the y-axis; see Figure 5. Based on Equations (3)–(5), measurement of the polarization dependence of the SHG signal hence reveals the orientation of the ferroelectric polarization. In the general case, additional aspects increase the complexity of Equations (3)–(5). For example, in the case of a multidomain pattern, the spontaneous polarization of the single domain is either replaced by the net polarization of the domain pattern or the respective SHG waves associated to each contributing domain state are added up to reveal the total SHG yield of the domain pattern [88]. Furthermore, surface SHG contributions and Fresnel coefficients in the beam propagation [89] may have to be considered.

2.2. Special Considerations in Thin Film SHG Experiments

Nonexistent in bulk systems, the presence of a substrate in the thin film system can have a significant influence on the SHG response. The substrate material may itself produce a non-zero SHG signal (see Figure 7) [88] that can interfere with the thin film response. Additionally, due to the breaking of inversion symmetry at surfaces and interfaces, there is often a surface contribution to the SHG response [86,88,90]. These additional components should be considered in thin film experiments as they may be a dominant part of the SHG signal in reduced volume conditions of the ferroic layer.

On the other hand, such parasitic SHG provides a reference light wave which can be used in interference experiments to achieve domain contrast in imaging experiments [79,86,88]. The substrate-induced SHG response in the case of a DSO substrate measured in transmission is shown in Figure 7a. The surface-related SHG signal emerging from a centrosymmetric $SrRuO_3$ (SRO) metallic oxide measured in reflection is shown in Figure 7b.

For full access to the possible orientations of polar axes in a thin film material, it is often desirable to generate SHG in transmission where surface contributions can be avoided in normal incidence configuration. However, while thin films of only a few tens of nanometers in thickness have negligible absorption throughout the visible and near-infrared spectrum, the absorption spectrum and possible birefringence effects of some substrates might affect the thin film SHG detection [16,80]. To minimize the substrate SHG contributions in thin film measurements, it may be advantageous to use an experimental setup in reflection geometry. There is therefore no standard SHG setup for thin films analysis and the balance between substrate, surface, and thin film contribution must be optimized in each measurement.

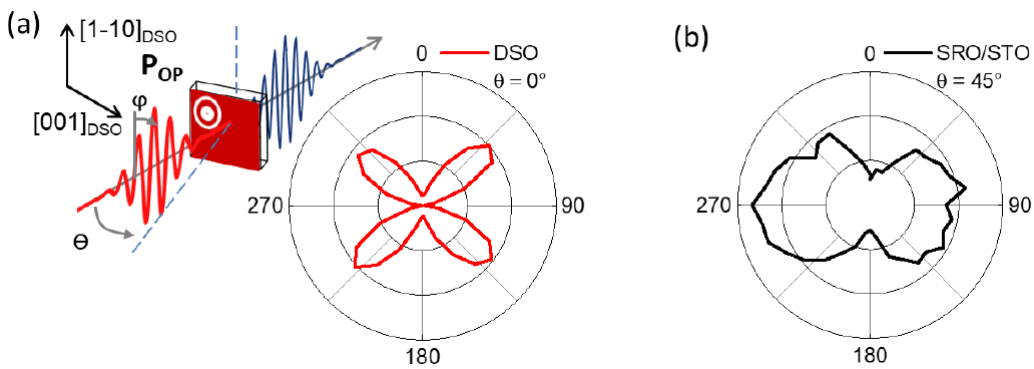


Figure 7. Substrate- and surface-related SHG. (**a**) Polarizer measurements performed on DSO in normal incidence (0° tilt angle) transmission. (**b**) Surface-related SHG response in anisotropy configuration at 45° incidence in reflection from an SRO thin film grown on STO substrate.

Finally, in contrast to many bulk systems, the expected domain sizes in thin films are usually below the optical SHG resolution limit. While certain sub-resolution features can be detected using SHG under the proper conditions, the efficiency to probe a ferroelectric state in thin films with SHG truly comes into its own when single domain or macroscopically-ordered domain systems are selected, i.e., when a macroscopic net polarization is present.

3. Probing Ferroic States in Thin Films Using SHG

In the following sections we review the major areas of application of SHG for probing ferroic states in thin film single layers and heterostructures. Results dealing with prototypical ferroelectric $PbZr_{0.2}Ti_{0.8}O_3$ (PZT) and multiferroic BFO materials are highlighted.

3.1. Optical Second Harmonic in Domain Engineered Ferroic Thin Films

The SHG technique is a powerful technique for the analysis of polar states in thin films possessing a net ferroelectric polarization, i.e., in a single domain state or in domain-engineered systems [88]. Uniaxial tetragonal ferroelectric $PbZr_{0.2}Ti_{0.8}O_3$ (PZT), the most technologically relevant ferroelectric compound, shows a thickness-dependent domain distribution [91]. A fully strained film can exhibit a single domain state with a polar axis pointing out of plane, along the long tetragonal *c*-axis. Above a threshold thickness (of about 60 nm in the case of PZT grown on $LaAlO_3$ substrates [91]), strain relaxation will induce a coexistence of nanosized in-plane oriented domains, i.e., *a*-domains with a polarization pointing along the in-plane oriented long axis, together with the out-of-plane *c*-domains [92,93].

The *a*-domains nucleate from either the top or the lower interfaces and their distribution in the volume can have major implication on the ferroelectric switching dynamics and thus on the targeted applications [94–96]. It is, however, non-trivial to access the ferroelectric domain distribution in the thin film volume when using scanning probe techniques due to their limited depth sensitivity.

As discussed in the previous section, the optical SHG technique only probes polar distortions projected on a plane perpendicular to the k-vector of the incoming light wave [16–18]. It is, therefore, possible to disentangle the contribution of *a*- and *c*-domains by changing the orientation of the ferroelectric polar axis with respect to the incident light beam [97].

The single-domain state of a thin film (15 nm thick) with a polarization pointing out of plane can be confirmed by PFM as shown in Figure 8a. For a fully *c*-oriented tetragonal film with point group 4*mm*, three independent χ_{ijk} components (χ_{zzz}, χ_{zxx} and χ_{xzx}) are non-zero [80,97]. At normal incidence ($\theta = 0°$), no second harmonic response is expected since the out-of-plane directed ferroelectric polarization then points along the k-vector. When tilting the ferroelectric system with respect to the incident light beam, the increment of the incidence angle will result in an increase of the associated SHG signal as shown in Figure 8b. The specific light polarization configuration is described in the inset of Figure 8b. The angular dependence of the SHG light in the polarizer measurement identifies the projection of the polar distortion to be along the direction of the double-lobe shape, as seen in Figure 8c. Here, the $[001]_{DSO}$ direction of the substrate is orthogonal to the tilt-axis ($[1–10]_{DSO}$) and represents the projection of an out-of-plane polarization.

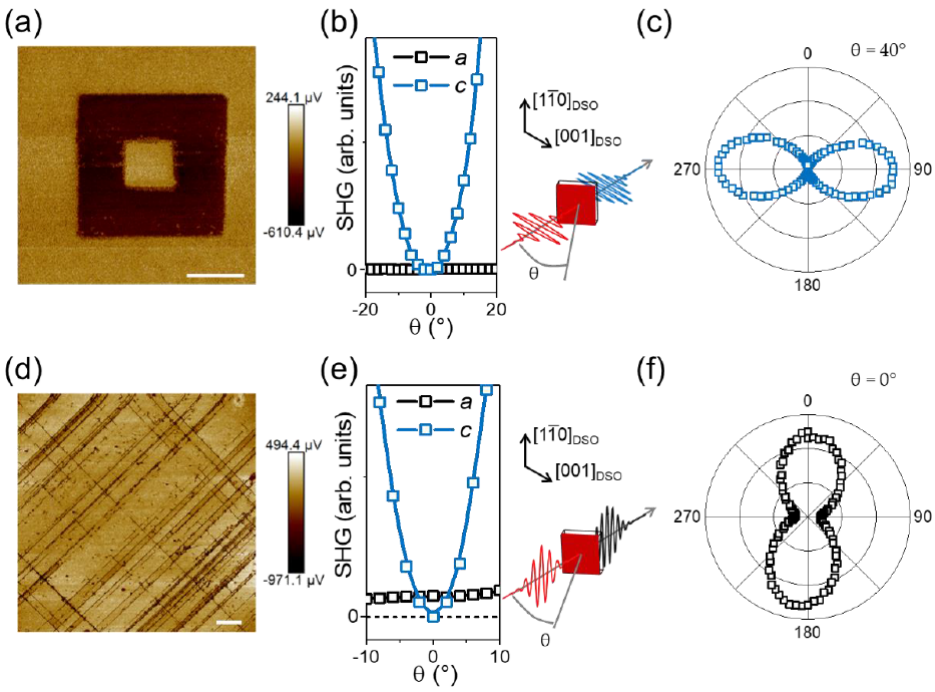


Figure 8. Optical detection of *a*-domains in PZT. (**a**) PFM scan of a single domain, *c*-oriented film. The scale bar is 1 μm. (**b**) SHG tilt scan showing the increased contribution from *c*-domains with increased tilting angle. The scale bar is 2 μm. (**c**) Polarizer measurement at a 40° tilt angle evidencing a single polar axis in the film. (**d**–**f**) Corresponding PFM, tilt scan and polarizer measurement at normal incidence for a mixed *a*-/*c*-domain state, respectively.

For a thicker layer, the epitaxial strain relaxation promotes the growth of nanosized stripe-domains with an in-plane polarization orientation, as evidenced by the PFM scan performed on the 75 nm thick film in Figure 8d. Here, specific light polarization configurations allow discriminating between the in-plane (in black) and the out-of-plane (in blue) contributions to the nonlinear optical response, as shown in Figure 8e. It is worth emphasizing that the SHG intensity associated with the in-plane

domains is non-vanishing at normal incidence ($\theta = 0°$), where no contribution is expected from *c*-axis (out-of-plane) domains. The polarizer measurement shown in Figure 8f highlights the presence of a macroscopic in-plane polarization directed along $[1\text{–}10]_{DSO}$. Note that in the presence of a complex sub-optical-resolution domain architecture where two sets of orthogonally oriented *a*-domains coexist [93], the SHG signal will not yield the spontaneous polarization of each domain but rather reflect the direction of the macroscopic in-plane net polarization resulting from the sum of all the *a*-domain contributions. In the present case, this corresponds to a 45° rotation of the polar axis shown in Figure 8f, as further discussed below for the case of BFO films. SHG is therefore sensitive to ferroelectric domains smaller than the optical resolution and the presence of *a*-domains throughout the volume of the film can be identified optically [97]. We note however that complementary scanning transmission electron microscopy is necessary to address specific domain morphologies throughout the film thickness [97].

The correspondence between the angular dependence of SHG and the orientation of the polar axis can be exploited further to identify the direction of a net polarization when more complex domain structures are considered [88]. For example, domain-engineered multiferroic BFO films subject to an in-plane anisotropic strain present two ferroelectric variants arranged in a stripe pattern (Figure 9a) [88]. The in-plane projections of the polarization in adjacent domains differ by 90° as highlighted by the small white arrows in the PFM scan in Figure 9a. This produces an average net in-plane polarization orthogonal to the domain wall elongation (indicated by the large white arrow in Figure 9a). The angular dependence of the SHG light therefore manifests as a two-fold shape oriented along the macroscopic polar axis direction as shown in Figure 9b.

In the absence of such in-plane anisotropy, four equivalent ferroelectric variants nucleate, characterized by the absence of an average macroscopic in-plane polarization (Figure 9c). In this case, the associated SHG signal does not depend on the sample orientation; see Figure 9d.

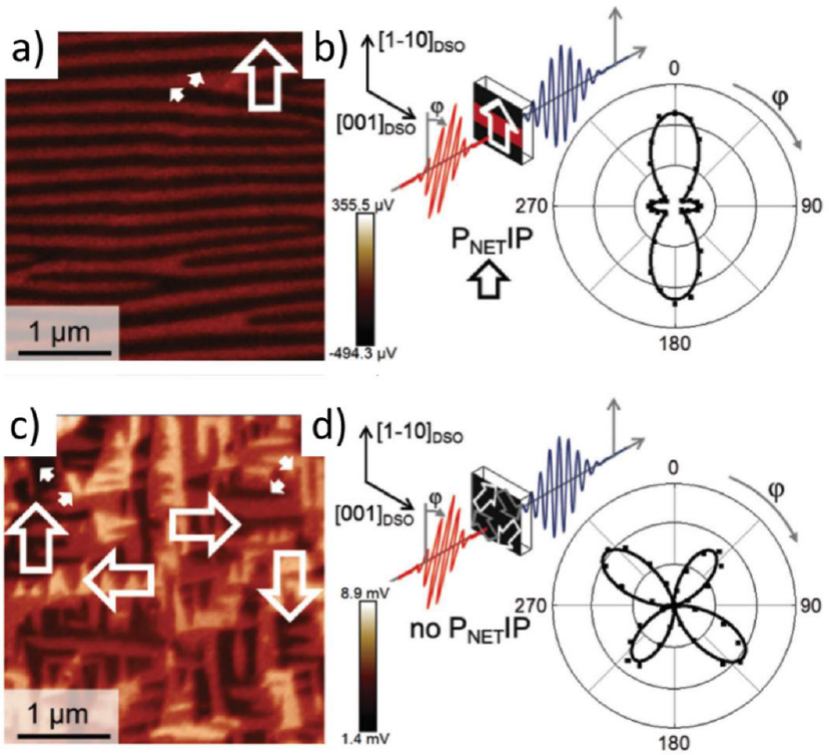


Figure 9. SHG on domain-engineered multiferroic BFO. (**a**) PFM scan of a stripe-like domain BFO film. (**b**) Polarizer measurement evidencing a net in-plane polarization in the film. (**c**,**d**) Corresponding PFM and polarizer measurement for an isotropic domain distribution, respectively. The signal observation in (**d**) does not depend on the sample crystalline orientation. Reproduced with permission from [88], Copyright Wiley-VCH Verlag GmbH and Co. KGaA, 2015.

SHG can also detect subtle structural changes such as strain dependent monoclinic distortions in epitaxial films. In particular, large compressive strains (>4.5%) in BFO films induce a tetragonal-like distortion favoring the emergence of a pseudo-tetragonal phase, i.e., BFO T-phase, while smaller compressive strains are associated with a rhombohedral-like distortion stabilizing the pseudo-rhombohedral phase, i.e., BFO R-phase [98,99]. These two phases possess the same monoclinic point group *m* but are characterized by a rotation of the in-plane polarization of 45° at the unit-cell level [100]. With SHG, these two phases can be simply distinguished by observing the direction of the double-sided lobe [88,101–103] as shown in Figure 10a,b. Note that the orientation of the polar axis is inclined by a few degrees with respect to the crystallographic axes due to the monoclinic distortion.

The R- and T-phases can also coexist in the same thin film system [98]. At the strain-driven morphotropic phase boundary, exotic conducting states and enhanced piezoelectric coefficients have been observed [104,105]. The SHG contributions associated with the two phases have different relative intensities as a function of the angle of incidence [101–103,106] as shown in Figure 10c. The pseudo-tetragonal phase (in red) has a symmetric behavior similar to a genuine tetragonal structure (cf. Figure 8b), while the SHG response of the R-phase (in green) reflects its polarization anisotropy. These distinct behaviors allow associating the nonlinear optical output at different incidence angles to the two ferroelectric phases, as shown in Figure 10d,e. In particular, a measurement at $\theta = 45°$ will be sensitive to the T-phase, while at normal incidence the SHG output is associated to the R-phase.

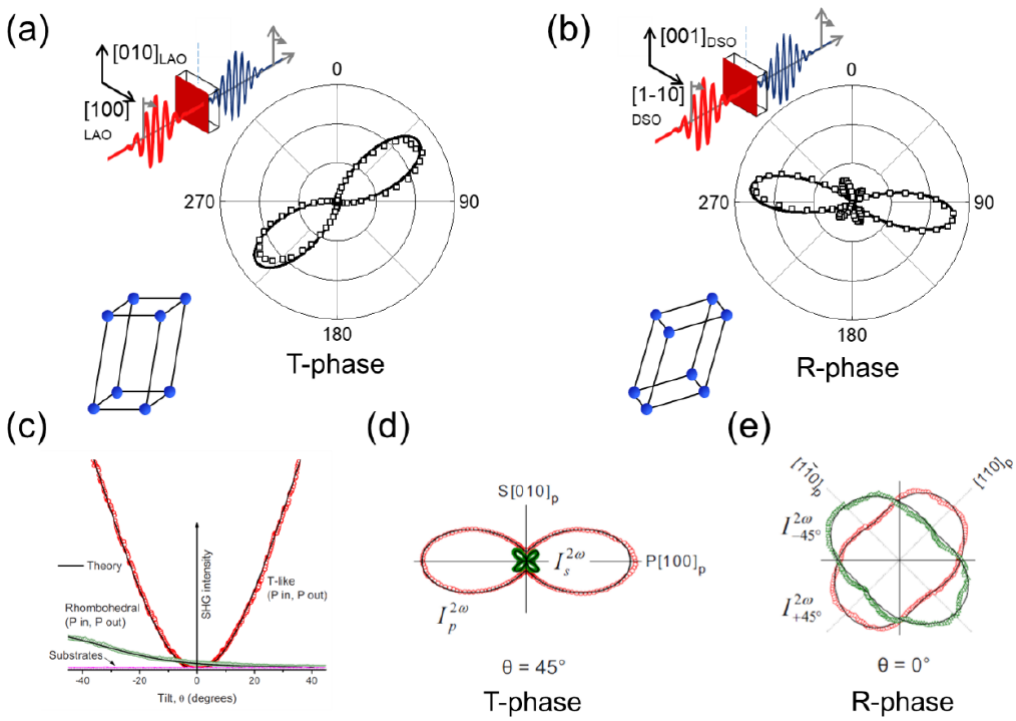


Figure 10. SHG to distinguish different monoclinic phases. (**a**,**b**) SHG anisotropy measurements collected at normal incidence for T- and R-phase, respectively. Reproduced with permission from [88] Copyright Wiley-VCH Verlag GmbH and Co. KGaA, 2015. (**c**) SHG intensity for different BFO films at different angles of incidence. (**d**,**e**) Polarizer measurements allow distinguishing the two phases. (**c**–**e**) Reprinted with the permission from [101], Copyright AIP Publishing, 2010.

For a fixed substrate, i.e., strain state, the strain relaxation with increasing thickness will modify the relative ratio of BFO T- and R-phases. This structural evolution results in a symmetry change for BFO films at various thicknesses that can be studied with SHG [107]. In ultrathin films grown in a compressive strain state using Nb-doped STO as substrate, a four-fold rotational symmetry associated with the absence of in-plane anisotropy is attributed to the T-phase [107] as shown in Figure 11a.

Around 100 nm, there is a transition regime where both phases coexist, while at 180 nm the polar axis of the R-phase is fully developed as observed in Figure 11b,c.

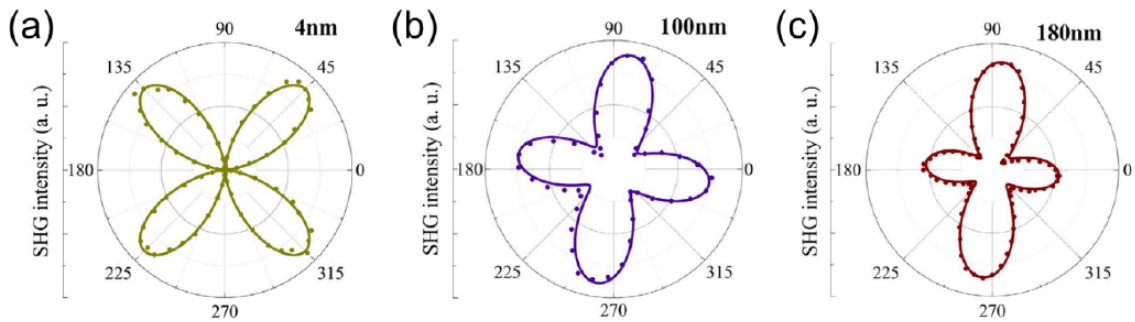


Figure 11. SHG to probe strain relaxation. (**a**–**c**) Polarizer measurements for different BFO film thicknesses can quantify the amount of T- and R-phases associated with a certain thickness. Reprinted from [107].

3.2. Imaging of Ferroic Switching Events

Optical SHG polarizer and anisotropy measurements probe the net polarization state of thin films in various domain configuration and phases. To visualize domains, domain walls and electric field induced switching events, the symmetry investigation can be combined with spatial resolution when performing SHG imaging.

Understanding the dynamics and switching mechanisms of ferroic switching events is the first step towards a deterministic control over the targeted functionality. While ferroelectric switching events can be probed at the nanoscale using a combination of STEM and PFM [95–97], complex switching events in magnetoelectric multiferroic compounds require the ability to map several coexisting order parameters at the single domain scale. In BFO thin films, the electric-field-induced change in the antiferromagnetic domain configuration has been demonstrated [108,109] at the scale of the ferroelectric domains. In Figure 12a,b, the local change of magnetic order after voltage application is shown using a combination of PFM and photoemission electron microscopy (PEEM) [108]. This seminal work established the one-to-one coupling between the ferroelectric domain architecture and the corresponding antiferromagnetic domains. The electric field control over this domain correlation led to the deterministic observation of electric field induced magnetization reversal in BFO-based heterostructure at room temperature [11–14,110].

Using spatially-resolved SHG, the respective domain structure of each ferroic order and their mutual coupling can be non-invasively probed. As mentioned previously, the as-grown domain architecture in ferroic thin films is typically well below the optical resolution limit (see Figure 12). On the other hand, using scanning probe techniques, it is possible to electrically write macroscopic single domains that are large enough for SHG imaging. Taking advantage of interference effects between oppositely polarized domains and surface contributions [88], one can selectively resolve either the voltage induced domain walls or domains; see Figure 12c–e. Starting from an as-grown ordered ferroelectric domain pattern, SHG imaging together with symmetry component analysis can distinguish between different types of scanning tip-induced switching events [88]. Therefore, specific switching events induced by the trailing field (originating from the lateral movement of the tip during the scan) [111,112] can be optically probed, as shown in Figure 12f,g.

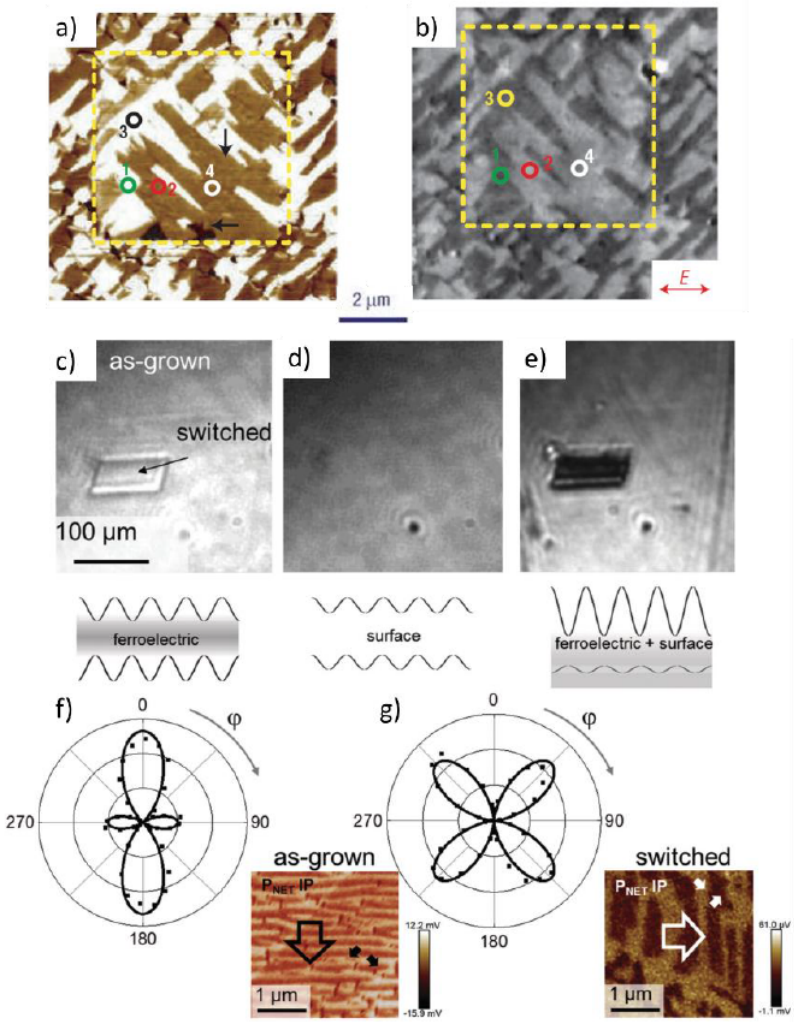


Figure 12. Switching events in BFO thin films. (**a**,**b**) Coupled switching of ferroelectric, (**a**) and antiferromagnetic (**b**) orders in BFO thin films. The one-to-one domain correlation persists after local voltage application (within the area delimited by the yellow dashed line). Reprinted by permission from [108], Copyright Macmillan Publishers Ltd.: Nature Materials, 2006. (**c**–**e**) SHG images of an SPM tip-poled 100 × 50 µm^2 field in BFO on DSO (110), using surface SHG interference to selectively image domain walls (**c**), surface (**d**) or domain polarization direction (**e**), respectively. (**f**,**g**) Corresponding SHG polarization dependence and in-plane PFM images outside (**f**) and inside (**g**) the switched area (**c**–**g**). Reproduced with permission from [88], Copyright Wiley-VCH Verlag GmbH and Co. KGaA, 2015.

Furthermore, the simultaneous evolution of several ferroic orders upon application of external fields can also be visualized using spatially resolved SHG [113]. By symmetry component analysis of the local SHG response, contributions from both ferroelectric domains and antiferromagnetic domain variants were deconvoluted in BFO thin films (Figure 13). The individual symmetry components were isolated using spatially resolved polarizer measurements. The anisotropy of the light polarimetry distribution was identified and correlated to the presence of different antiferromagnetic domain variants, superimposed on the ferroelectric response; see Figure 13b. Comparing the spatially resolved SHG response of the domain structure before and after application of a scanning probe-tip induced external electric field, it could be observed that the switching of the ferroelectric polarization was accompanied by a change in the antiferromagnetic domain state [113]. This optical ferroic domain-mapping capacity demonstrated in the case of BFO thin films opens new avenues toward the determination of magnetoelectric switching dynamics in multiferroic-based heterostructures.

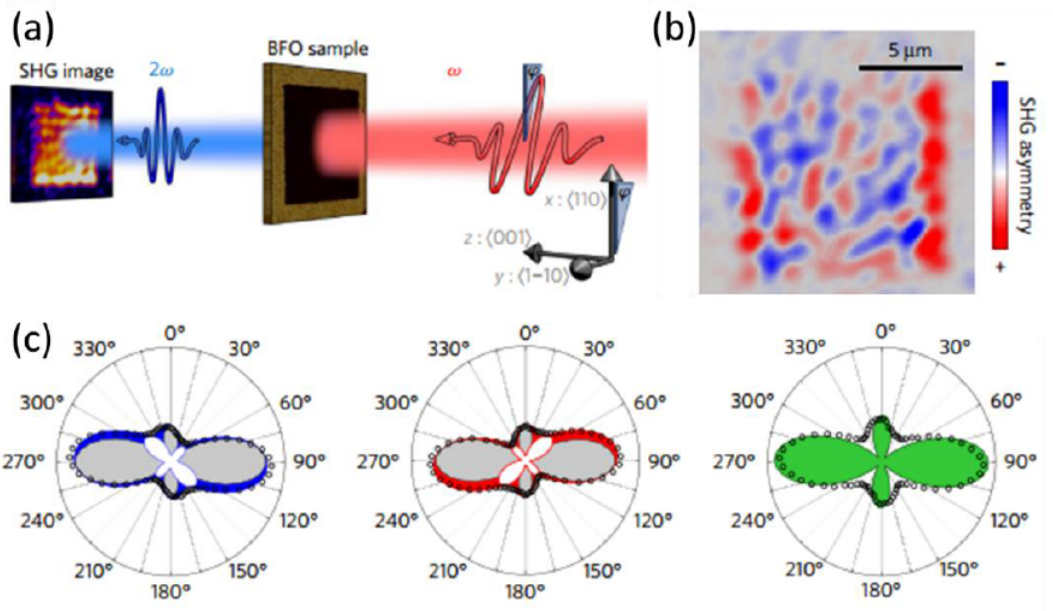


Figure 13. Multiferroic domain mapping in a BFO thin film. (**a**) SHG imaging set up. (**b**) The spatially resolved anisotropy corresponding to antiferromagnetic domains. (**c**) Polarizer measurements in different areas of the sample showing a spatially distributed anisotropy in the polarizer symmetry. Reprinted by permission from [113]. Copyright Macmillan Publishers Ltd.: Nature Materials, 2017.

It has to be emphasized here, that the polarization orientation in thin films can be probed with SHG even when the layer is capped with a top electrode [88]. Figure 14 shows the PFM and SHG response of a BFO thin film capped with 4 nm Pt. While no in-plane PFM domain contrast is observed, the polarizer SHG measurement can identify the direction of the macroscopic polar axis of the buried BFO layer (see Figure 9b). Contrary to SPM techniques, the SHG signal can therefore be used to probe a polar state in a buried layer and is therefore appropriate for in-operando monitoring of multiferroic magnetoelectric switching events and ferroic domain dynamics.

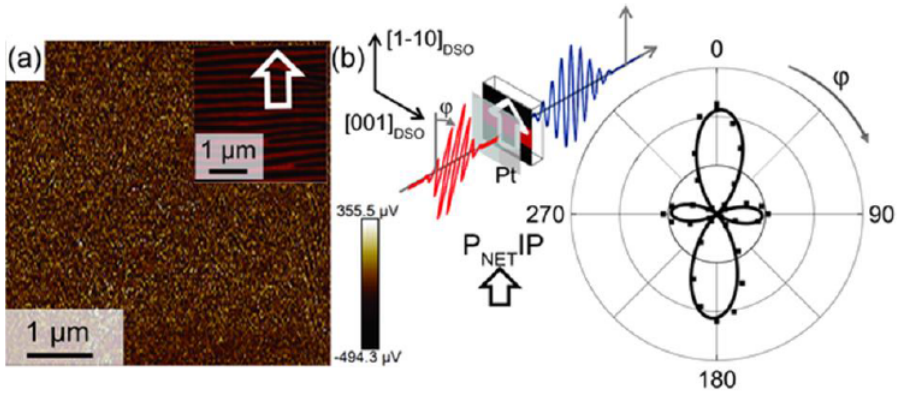


Figure 14. Ferroelectric domain structure in a BFO thin film on $La_{0.7}Sr_{0.3}MnO_3$ buffered DSO capped with 4 nm Pt. (**a**) PFM image of the Pt-capped BFO shows no ferroelectric domain contrast. For comparison, the inset shows the domain structure of the uncapped BFO layer. (**b**) SHG polarizer measurement identifies a polar axis along the vertical orientation, corresponding to the macroscopic net polarization direction, indicated by the white arrows in the insets of (**a**,**b**). Reproduced with permission from [88], Copyright Wiley-VCH Verlag GmbH and Co. KGaA, 2015.

3.3. Domain Wall Architecture

During the last decade, a wide variety of novel phenomena has been discovered at domain walls in ferroic systems ranging from unusual electronic conductance to inseparably entangled spin and charge degrees of freedom [114]. Ferroic domain walls are two-dimensional functional objects that can be created and erased on demand, introducing new opportunities for the next generation of rewritable nanoelectronics [115–117].

To investigate these nanosized objects, a combination of multiple techniques is required. For instance, a conducting scanning probe tip can be used to write a desired domain pattern and create 180° domain walls, but it cannot provide information on their internal architecture, i.e., polarization orientation, symmetry, etc. STEM is commonly used to investigate the atomic configuration across a domain wall section [97,118] but the information is limited to a single projection perpendicular to the domain wall propagation. SHG can provide complementary information on the polar structure of the domain walls. By directly relating the nonlinear susceptibility tensor elements with the symmetry linked to ferroic order in the system, SHG can selectively pick up the signal from different types of domain-wall configurations, determining their internal polarization orientation [119]. The domain wall architecture can therefore be investigated in three dimensions using a combination of STEM and SHG for cross-section and top view structure analysis, respectively.

The SHG image of a ferroelectric PZT thin film exhibiting a mixed *a*-/*c*-domain state is shown in Figure 15a. The image, acquired with an in-plane polarization sensitivity [97], displays stripes along the $[001]_{DSO}$ orientation corresponding to $[1–10]_{DSO}$-polarized *a*-domains together with an artificially poled region.

Strikingly, a non-zero SHG intensity is observed at the edges of the tip-poled region which cannot be explained by the assumed Ising-like 180° domain walls architecture in strained thin films [97,120]. At 180° Ising-like domain walls, the contributions of SHG response from the neighboring opposite domains interfere destructively due to a 180° phase difference [79,88]. In addition, in the present optical configuration, i.e., normal incidence, the sensitivity is not set to an out-of-plane oriented polarization. The likely explanation for a non-zero SHG signal at the domain wall is therefore the existence of an in-plane polarization component within the ferroelectric domain wall itself. This corresponds to the presence of a Néel or Bloch component of the polarization, usually observed at domain walls in ferromagnetic systems [121]. It is also interesting to note that, although the PFM line profile in the switched region in Figure 15b shows a single *c*-domain state, the SHG signal profile, which probes the total film thickness, reveals the presence of persisting buried *a*-domains in the same region [97].

The combination of STEM and SHG allows the investigation of 180° domain walls in cross-section and top view, respectively. The Ising component observed when mapping atomic displacement across the wall (Figure 15c) combined with a polarization component perpendicular to the domain wall plane identified by SHG imaging (Figure 15a), led to the demonstration of mixed Ising-Néel behavior of the PZT thin film ferroelectric wall [97,122].

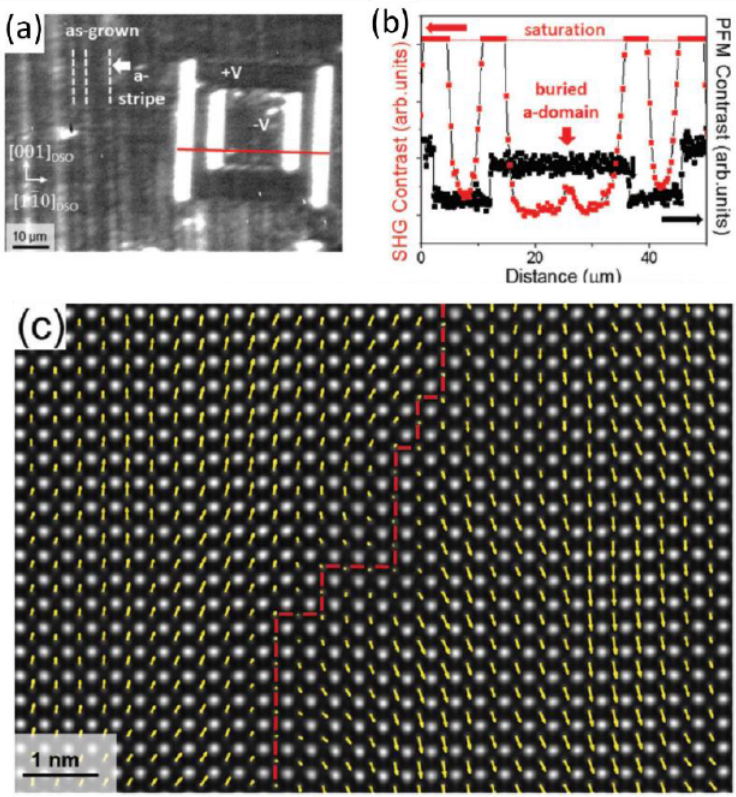


Figure 15. Polar behavior of a ferroelectric domain wall. (**a**) SHG imaging a stripe *a*-domain pattern in a PZT thin film. A tip-induced switching event is visible on the right side (+V and –V indicate where a positive and negative voltage was applied by the scanning tip, respectively). (**b**) Piezoresponse force microscopy and SHG intensity line scans over the section indicated by the red line in (**a**). (**c**) High-angle annular dark-field (HAADF) STEM image overlaid with polarization vectors showing a smooth Néel-like rotation combined with an Ising-like polarization decrease across the domain wall. Reproduced with permission from [97], Copyright Wiley-VCH Verlag GmbH and Co. KGaA, 2016.

4. Future Perspectives

4.1. Remaining Challenges in 3D Resolution in Thin Films

In standard ferroelectrics, 180° domain walls used to be considered Ising-like due to the strong correlation between strain and polar distortion [123]. The observation of tilted mixed Ising-Néel walls in thin films is suggesting new opportunities at these functional interfaces—one can push the analogy further, to magnetic domain walls, and expect to probe a Bloch behavior at a ferroelectric thin film domain wall. In $LiTaO_3$ crystals [122], the detection of a polarization component within the plane of the wall by SHG allowed the identification of the Bloch-like nature of 180° domain walls, as shown in Figure 16a. The observation of such a wall type in thin films is still unreported.

In bulk materials, imaging techniques are gaining increasing interest with the push towards the internal structure analysis of single domain walls. The full mapping of ferroelectric domain walls can now be performed during their creation at the ferroelectric phase transition and in three dimensions using temperature dependent confocal SHG microscopy [124–126], as shown in Figure 16b in the case of ferroelectric triglycine sulfate (TGS) [127].

Regarding thin films, the optical investigation of domain walls remains challenging due to the optical resolution limit and the limited SHG active domain wall volume. Recent progress dealing with laser-scanning SHG microscopy [128] and near-field SHG nano-imaging [129,130] where tip-enhanced

SHG signals are collected with a spatial resolution approaching the submicron range [131,132] has been achieved and opens new avenues towards the symmetry analysis of nanoscale objects.

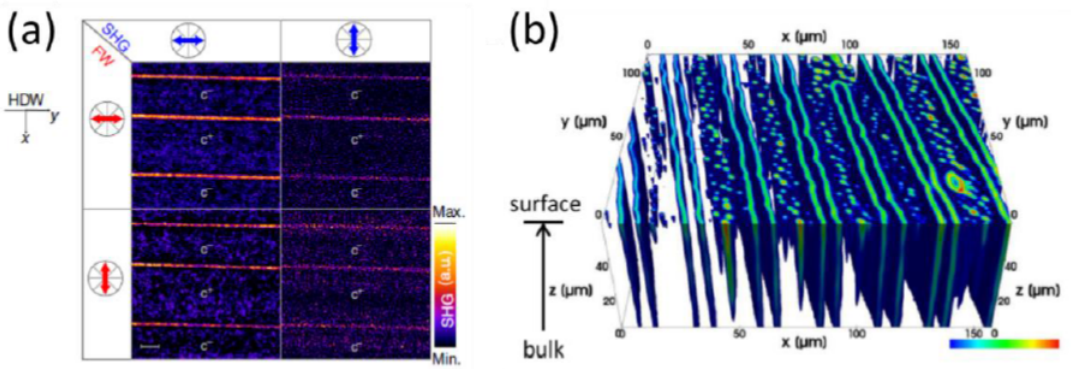


Figure 16. Three-dimensional structure of domain walls in ferroelectric crystals. (**a**) SHG image of periodic Bloch domain walls in $LiTaO_3$ crystals. Reprinted from [122]. (**b**) Three dimensional SHG microscopy image of domains walls in TGS. Reproduced with permission from [127], Copyright Wiley-VCH Verlag GmbH and Co. KGaA, 2017.

4.2. Probing Ferroelectric Heterostructures during the Growth Process

Significant advances have been made in the growth quality of thin film oxide ferroelectrics. Still, the process has been lacking control of the polar state in multilayer architectures. Most of the technological applications for ferroic thin films require a deterministic control of the domain state. Ferroelectric tunnel junctions require the stabilization of a single-domain state to maximize the polarization-orientation dependent change in conduction between the switched states [133,134]. In the case of tetragonal ferroelectric-based memory, the presence of the as-grown in-plane domains hinders the switching process of the out-of-plane polarization state [2,96]. Eliminating the in-plane domains in these systems would thus be of interest [135]. On the other hand, in multiferroic heterostructures aiming towards the electrical control of magnetization, a specific domain configuration triggers deterministic switching events [111].

To achieve full polarization control of a heterostructure, one needs to be able to monitor the polarization during the growth process. Most of the classical ferroelectric epitaxial strained systems are in fact grown in their polar phase [33,61,136]. However, none of the traditional thin-film characterization techniques can be implemented in situ due to their invasive nature and the extreme conditions during deposition, such as high substrate temperatures and pressure variations.

Synchrotron *x*-ray-based techniques, which are sensitive to periodicity in polar systems, i.e., lattice parameter and periodic domain splitting, have been used to show that past a critical thickness, BTO thin films experience an increased tetragonality during the growth process [21,136]. A schematic of the experimental setup is shown in Figure 17a,b with the corresponding measurements. These pioneering experiments revealed the emergence of the polar state during the growth and broadened the investigation of ferroelectric materials in situ. This technique, however, requires high intensity flux and is limited to highly periodic patterns.

SHG is sensitive to symmetry breaking at the emergence of all ferroic orders, which makes it an ideal technique for studying not only single- and complex-domain states but also the transition from one to another. Its non-invasive and non-destructive nature renders the optical coupling into a growth chamber conceivable [61,137–139]. Aside from a few attempts to characterize growth of metallic ferromagnets [138,140] and its use as a surface reaction probe [141], the SHG potential as oxide growth diagnostic tool during the pulsed laser deposition process has only recently been demonstrated [61]. A schematic of the in situ SHG (ISHG) experimental setup is shown in Figure 17c.

Direct access to the thin film polarization during its growth is achieved in real time using ISHG. The emergence of the ferroelectric polarization in BTO is shown in Figure 17d. The ability to probe multilayer systems [61,120] opens new avenues towards understanding the complex interplay between strain, interfaces, and electrostatics in oxide superlattices. Interfering signals originating from each polar constituent of a superlattice result in destructive or constructive interference depending on their relative polarization orientation. For instance, Figure 17e shows a succession of increasing ISHG intensity during the growth of ferroelectric BFO layers with parallel polarization orientation. The growth of oppositely polarized layers results in a decrease of the signal [61].

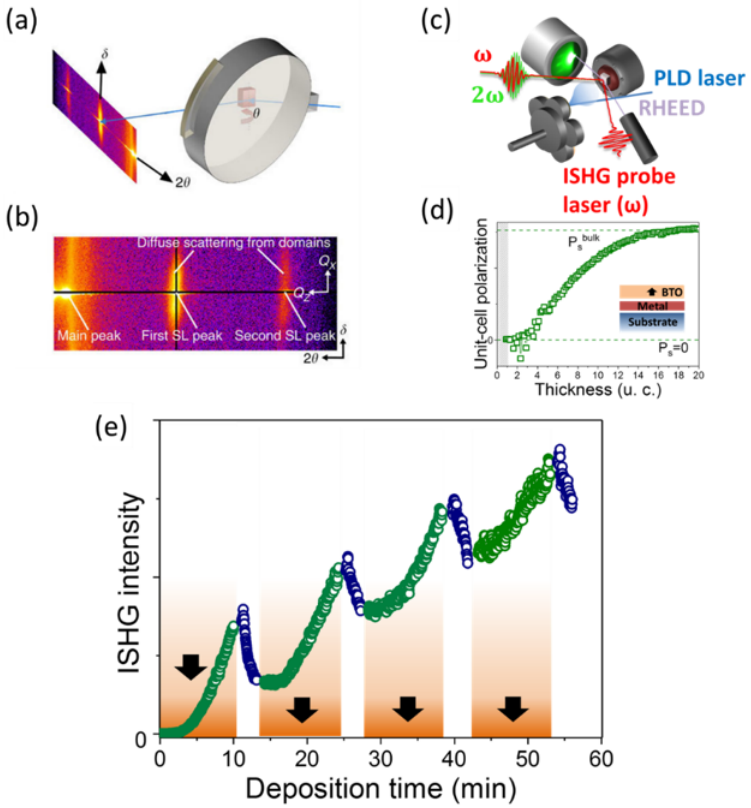


Figure 17. Characterization of the polar state during the growth. (**a**) Schematic of the in situ X-ray diffraction experiment, (**b**) Example of diffraction results for a BTO/STO superlattice. (**a**,**b**) Reprinted from [21]. (**c**) Schematic of the in situ SHG experimental setup. (**d**) In situ SHG measurement showing the emergence of polarization during the growth of a BTO thin film. (**e**) In situ SHG during the deposition of a BFO/SRO multilayer with a parallel polarization configuration. (**c**–**e**) Reprinted with from [61].

Depolarizing field, strain, interfaces, and thickness-related effects on the ferroic state of oxide thin films can now be accessed in real time during the synthesis being the key to full deposition and functionality control for interface-based oxide-electronic devices.

4.3. Towards Magnetoelectric Switching Dynamics

Now that non-destructive access to a buried ferroelectric domain configuration is gained using SHG, tracking its dynamic response to external perturbations is enabled. It is possible, for example, to apply an electric field to magnetoelectric devices and track the resulting shifts in the magnetic domain pattern. To the best of our knowledge, only a few experiments were carried out with the scope of monitoring the multiferroic, respectively magnetoelectric switching dynamics [13,142,143]. In the case of BFO thin films, the kinetics of the electric field induced reversal of the canting of the antiferromagnetically aligned spins has been investigated using time dependent dual-frequency PFM

studies and first principle calculations [13]. A two-step switching process of the coupled polarization and the antiferromagnetic order results in the observed electric field induced magnetization reversal.

In another study, SHG measurements performed on bulk $MnWO_4$ revealed a surprisingly slow reversal of the magnetic order parameter by an electric field on the time scale of milliseconds [143]. Hence, although understanding the *switching dynamics* rather than solely the *presence* of a magnetoelectric coupling is essential for any type of device application, research in this direction remains an underdeveloped topic. The non-invasive nature of SHG may now pave the way to such investigations [88,113] in multifunctional thin film architectures.

Parallel to this, alternatives to multiferroic and magnetoelectric switching mechanisms are gaining increasing interest in the push for low-energy-consuming electric-field control of magnetization. Spin-orbit-torque-based magnetic domain switching may serve as an example here [144]. Recent experiments have demonstrated a current induced magnetic switching in the picosecond regime [145].

5. Conclusions

Optical SHG is an established tool to investigate the ferroic nature of single crystals. It enabled major achievements in the field of ferroelectrics and multiferroics [16–18]. It relies on the ability to selectively probe several order parameters in one experiment which has a great impact on multiferroic dynamics [38] and on the observation of coupled magnetic and electric domains in a multiferroic compound [146]. Recent advances in ferroic domain engineering in thin films using epitaxial strain as a new degree of freedom are progressively rendering optical characterization of nanoscale ordered ferroic domains in thin films possible. Polar axis orientation, Curie temperature, strain tuning of the ferroelectric polarization and electric field induced switching events can now be probed non-invasively with monolayer accuracy.

Because the SHG probe has a long working distance, the investigation of the emergence of the polarization during the thin film deposition process can be designed [61]. In parallel with in situ structural knowledge gained with RHEED, in situ SHG gives the functional information in real time during the synthesis. The ISHG characterization is not affected by the extreme conditions during the growth process and can improve the understanding of ferroelectricity from the first unit cell deposited. The direct access to the ferroic domain state during growth is opening new routes towards the full control over ferroelectricity. The manipulation of the polarization orientation during the growth in a multilayer will have an important role in future applications. In addition, artificial domain walls with arbitrary polarization sequences [120] offer a wide range of physical properties [11,114,120] and can for instance be designed on demand. A possible application would deal with the ability to electrically tune the symmetry environment of a layer sandwiched between two ferroelectric films [120]. The capacity to set and verify buried polar states during the growth [61] allows tuning of the strain, charge and symmetry environment of an adjacent active layer such as symmetry sensitive topological states, metal insulator transitions or colossal magnetoresistance systems. Furthermore, in situ SHG will support efforts in the understanding of polar states in the ultrathin regime and in complex oxide-based heterostructures and superlattices. For example, the development of the technologically promising voltage amplification based on transient negative capacitance in ferroelectric heterostructures will greatly benefit from the investigation of the early stage of ferroelectric vortex formation [50,51].

Acknowledgments: The authors thank J. Schaab for fruitful discussions. The authors acknowledge funding through the SNSF R'Equip Program (Grant. No. 206021-144988).

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Scott, J.F.; De Araujo, C.A.P. Ferroelectric memories. *Science* **1989**, *246*, 1400–1405. [CrossRef] [PubMed]
2. Setter, N.; Damjanovic, D.; Eng, L.; Fox, G.; Gevorgian, S.; Hong, S.; Kingon, A.; Kohlstedt, H.; Park, N.Y.; Stephenson, G.B.; et al. Ferroelectric thin films: Review of materials, properties, and applications. *J. Appl. Phys.* **2006**, *100*, 051606. [CrossRef]
3. Garcia, V.; Bibes, M.; Bocher, L.; Valencia, S.; Kronast, F.; Crassous, A.; Moya, X.; Enouz-Vedrenne, S.; Gloter, A.; Imhoff, D.; et al. Ferroelectric Control of Spin Polarization. *Science* **2010**, *327*, 1106–1110. [CrossRef] [PubMed]
4. Chanthbouala, A.; Garcia, V.; Cherifi, R.O.; Bouzehouane, K.; Fusil, S.; Moya, X.; Xavier, S.; Yamada, H.; Deranlot, C.; Mathur, N.D.; et al. A ferroelectric memristor. *Nat. Mater.* **2012**, *11*, 860–864. [CrossRef]
5. Junquera, J.; Ghosez, P. Critical thickness for ferroelectricity in perovskite ultrathin films. *Nature* **2003**, *422*, 506–509. [CrossRef] [PubMed]
6. Fong, D.D.; Stephenson, G.B.; Streiffer, S.K.; Eastman, J.A.; Auciello, O.; Fuoss, P.H.; Thompson, C. Ferroelectricity in ultrathin perovskite films. *Science* **2004**, *304*, 1650–1653. [CrossRef] [PubMed]
7. Schlom, D.G.; Chen, L.Q.; Eom, C.B.; Rabe, K.M.; Streiffer, S.K.; Triscone, J.M. Strain tuning of ferroelectric thin films. *Annu. Rev. Mater. Res.* **2007**, *37*, 589–626. [CrossRef]
8. Damodaran, A.R.; Agar, J.C.; Pandya, S.; Chen, Z.; Dedon, L.; Xu, R.; Apgar, B.; Saremi, S.; Martin, L.W. New modalities of strain-control of ferroelectric thin films. *J. Phys. Condens. Matter* **2016**, *28*, 263001. [CrossRef] [PubMed]
9. Scott, J.F. Multiferroic memories. *Nat. Mater.* **2007**, *6*, 256–257. [CrossRef] [PubMed]
10. Gajek, M.; Bibes, M.; Fusil, S.; Bouzehouane, K.; Fontcuberta, J.; Barthélémy, A.; Fert, A. Tunnel junctions with multiferroic barriers. *Nat. Mater.* **2007**, *6*, 296–302. [CrossRef] [PubMed]
11. Trassin, M. Low energy consumption spintronics using multiferroic heterostructures. *J. Phys. Condens. Matter* **2016**, *28*, 033001. [CrossRef] [PubMed]
12. Heron, J.T.; Trassin, M.; Ashraf, K.; Gajek, M.; He, Q.; Yang, S.Y.; Nikonov, D.E.; Chu, Y.H.; Salahuddin, S.; Ramesh, R. Electric-Field-Induced Magnetization Reversal in a Ferromagnet-Multiferroic Heterostructure. *Phys. Rev. Lett.* **2011**, *107*, 217202. [CrossRef] [PubMed]
13. Heron, J.T.; Bosse, J.L.; He, Q.; Gao, Y.; Trassin, M.; Ye, L.; Clarkson, J.D.; Wang, C.; Liu, J.; Salahuddin, S.; et al. Deterministic switching of ferromagnetism at room temperature using an electric field. *Nature* **2014**, *516*, 370–373. [CrossRef] [PubMed]
14. Heron, J.T.; Schlom, D.G.; Ramesh, R. Electric field control of magnetism using $BiFeO_3$-based heterostructures. *Appl. Phys. Rev.* **2014**, *1*, 021303. [CrossRef]
15. Bibes, M.; Barthélémy, A. Multiferroics: Towards a magnetoelectric memory. *Nat. Mater.* **2008**, *7*, 425–426. [CrossRef] [PubMed]
16. Denev, S.A.; Lummen, T.T.A.; Barnes, E.; Kumar, A.; Gopalan, V. Probing Ferroelectrics Using Optical Second Harmonic Generation. *J. Am. Ceram. Soc.* **2011**, *94*, 2699–2727. [CrossRef]
17. Fiebig, M.; Pavlov, V.V.; Pisarev, R.V. Second-harmonic generation as a tool for studying electronic and magnetic structures of crystals: Review. *J. Opt. Soc. Am. B* **2005**, *22*, 96–118. [CrossRef]
18. Yokota, H.; Kaneshiro, J.; Uesu, Y. Optical Second Harmonic Generation Microscopy as a Tool of Material Diagnosis. *Phys. Res. Int.* **2012**, *2012*, 704634. [CrossRef]
19. Khestanova, E.; Dix, N.; Fina, I.; Scigaj, M.; Rebled, J.M.; Magén, C.; Estradé, S.; Peiró, F.; Herranz, G.; Fontcuberta, J.; et al. Untangling electrostatic and strain effects on the polarization of ferroelectric superlattices. *Adv. Funct. Mater.* **2016**, *26*, 6446–6453. [CrossRef]
20. Liu, G.; Chen, J.; Lichtensteiger, C.; Triscone, J.M.; Aguado-Puente, P.; Junquera, J.; Valanoor, N. Positive Effect of an Internal Depolarization Field in Ultrathin Epitaxial Ferroelectric Films. *Adv. Electron. Mater.* **2016**, *2*, 1500288. [CrossRef]
21. Bein, B.; Hsing, H.C.; Callori, S.J.; Sinsheimer, J.; Chinta, P.V.; Headrick, R.L.; Dawber, M. In situ X-ray diffraction and the evolution of polarization during the growth of ferroelectric superlattices. *Nat. Commun.* **2015**, *6*, 10136. [CrossRef] [PubMed]
22. Zubko, P.; Gariglio, S.; Gabay, M.; Ghosez, P.; Triscone, J.M. Interface Physics in Complex Oxide Heterostructures. *Annu. Rev. Condens. Matter Phys.* **2011**, *2*, 141–165. [CrossRef]

23. Stephenson, G.B.; Highland, M.J. Equilibrium and stability of polarization in ultrathin ferroelectric films with ionic surface compensation. *Phys. Rev. B* **2011**, *84*, 064107. [CrossRef]
24. Lichtensteiger, C.; Dawber, M.; Stucki, N.; Triscone, J.M.; Hoffman, J.; Yau, J.B.; Ahn, C.H.; Despont, L.; Aebi, P. Monodomain to polydomain transition in ferroelectric $PbTiO_3$ thin films with $La_{0.67}Sr_{0.33}MnO_3$ electrodes. *Appl. Phys. Lett.* **2007**, *90*, 052907. [CrossRef]
25. Dawber, M.; Chandra, P.; Littlewood, P.B.; Scott, J.F. Depolarization corrections to the coercive field in thin-film ferroelectrics. *J. Phys. Condens. Matter* **2003**, *15*, L393–L398. [CrossRef]
26. Streiffer, S.K.; Eastman, J.A.; Fong, D.D.; Thompson, C.; Munkholm, A.; Ramana Murty, M.V.; Auciello, O.; Bai, G.R.; Stephenson, G.B. Observation of nanoscale 180 degrees stripe domains in ferroelectric $PbTiO_3$ thin films. *Phys. Rev. Lett.* **2002**, *89*, 067601. [CrossRef] [PubMed]
27. Tenne, D.A.; Bruchhausen, A.; Lanzillotti-Kimura, N.D.; Fainstein, A.; Katiyar, R.S.; Cantarero, A.; Soukiassian, A.; Vaithyanathan, V.; Haeni, J.H.; Tian, W.; et al. Probing nanoscale ferroelectricity by ultraviolet Raman spectroscopy. *Science* **2006**, *313*, 1614–1616. [CrossRef] [PubMed]
28. Dawber, M.; Stucki, N.; Lichtensteiger, C.; Gariglio, S.; Ghosez, P.; Triscone, J.M. Tailoring the properties of artificially layered ferroelectric superlattices. *Adv. Mater.* **2007**, *19*, 4153–4159. [CrossRef]
29. Lee, H.N.; Christen, H.M.; Chisholm, M.F.; Rouleau, C.M.; Lowndes, D.H. Strong polarization enhancement in asymmetric three-component ferroelectric superlattices. *Nature* **2005**, *433*, 395–399. [CrossRef] [PubMed]
30. Scott, J.F. Ferroelectrics go bananas. *J. Phys. Condens. Matter* **2008**, *20*, 021001. [CrossRef]
31. Rijnders, G.J.H.M.; Koster, G.; Blank, D.H.A.; Rogalla, H. In situ monitoring during pulsed laser deposition of complex oxides using reflection high energy electron diffraction under high oxygen pressure. *Appl. Phys. Lett.* **1997**, *70*, 1888–1890. [CrossRef]
32. Ederer, C.; Spaldin, N.A. Effect of Epitaxial Strain on the Spontaneous Polarization of Thin Film Ferroelectrics. *Phys. Rev. Lett.* **2005**, *95*, 257601. [CrossRef] [PubMed]
33. Choi, K.J.; Biegalski, M.; Li, Y.L.; Sharan, A.; Schubert, J.; Uecker, R.; Reiche, P.; Chen, Y.B.; Pan, X.Q.; Gopalan, V.; et al. Enhancement of Ferroelectricity in Strained $BaTiO_3$ Thin Films. *Science* **2004**, *5*, 1005–1009. [CrossRef] [PubMed]
34. Damodaran, A.R.; Breckenfeld, E.; Chen, Z.; Lee, S.; Martin, L.W. Enhancement of ferroelectric Curie temperature in $BaTiO_3$ films via straininduced defect dipole alignment. *Adv. Mater.* **2014**, *26*, 6341–6347. [CrossRef] [PubMed]
35. Xu, R.; Liu, S.; Grinberg, I.; Karthik, J.; Damodaran, A.R.; Rappe, A.M.; Martin, L.W. Ferroelectric polarization reversal via successive ferroelastic transitions. *Nat. Mater.* **2015**, *14*, 79–86. [CrossRef] [PubMed]
36. White, J.S.; Bator, M.; Hu, Y.; Luetkens, H.; Stahn, J.; Capelli, S.; Das, S.; Döbeli, M.; Lippert, T.; Malik, V.K.; et al. Strain-Induced Ferromagnetism in Antiferromagnetic $LuMnO_3$ Thin Films. *Phys. Rev. Lett.* **2013**, *111*, 037201. [CrossRef] [PubMed]
37. Becher, C.; Maurel, L.; Aschauer, U.; Lilienblum, M.; Magén, C.; Meier, D.; Langenberg, E.; Trassin, M.; Blasco, J.; Krug, I.P.; et al. Strain-induced coupling of electrical polarization and structural defects in $SrMnO_3$ films. *Nat. Nanotechnol.* **2015**, *10*, 661–665. [CrossRef] [PubMed]
38. Fiebig, M.; Lottermoser, T.; Meier, D.; Trassin, M. The evolution of multiferroics. *Nat. Rev. Mater.* **2016**, *1*, 16046. [CrossRef]
39. Fiebig, M. Revival of the magnetoelectric effect. *J. Phys. D Appl. Phys.* **2005**, *38*, R123–R152. [CrossRef]
40. Spaldin, N.A.; Fiebig, M. The Renaissance of Magnetoelectric Multiferroics. *Science* **2005**, *309*, 391–392. [CrossRef] [PubMed]
41. Wang, J.; Neaton, J.B.; Zheng, H.; Nagarajan, V.; Ogale, S.B.; Liu, B.; Viehland, D.; Vaithyanathan, V.; Schlom, D.G.; Waghmare, U.V.; et al. Epitaxial $BiFeO_3$ Multiferroic Thin Film Heterostructures. *Science* **2003**, *299*, 1719–1722. [CrossRef] [PubMed]
42. Martí, X.; Sánchez, F.; Hrabovsky, D.; Fàbrega, L.; Ruyter, A.; Fontcuberta, J.; Laukhin, V.; Skumryev, V.; García-Cuenca, M.V.; Ferrater, C.; et al. Exchange biasing and electric polarization with $YMnO_3$. *Appl. Phys. Lett.* **2006**, *89*, 032510. [CrossRef]
43. Laukhin, V.; Skumryev, V.; Martí, X.; Hrabovsky, D.; Sánchez, F.; García-Cuenca, M.V.; Ferrater, C.; Varela, M.; Lüders, U.; Bobo, J.F.; et al. Electric-Field Control of Exchange Bias in Multiferroic Epitaxial Heterostructures. *Phys. Rev. Lett.* **2006**, *97*, 227201. [CrossRef] [PubMed]
44. Ramesh, R.; Spaldin, N.A. Multiferroics: Progress and prospects in thin films. *Nat. Mater.* **2007**, *6*, 21–29. [CrossRef] [PubMed]

45. Shimamoto, K.; Mukherjee, S.; Manz, S.; White, J.S.; Trassin, M.; Kenzelmann, M.; Chapon, L.; Lippert, T.; Fiebig, M.; Schneider, C.W.; et al. Tuning the multiferroic mechanisms of $TbMnO_3$ by epitaxial strain. *Sci. Rep.* **2017**, *7*, 44753. [CrossRef] [PubMed]
46. Trassin, M.; Viart, N.; Versini, G.; Loison, J.L.; Vola, J.P.; Schmerber, G.; Crégut, O.; Barre, S.; Pourroy, G.; Lee, J.H.; et al. Epitaxial thin films of multiferroic $GaFeO_3$ on conducting indium tin oxide (001) buffered yttrium-stabilized zirconia (001) by pulsed laser deposition. *Appl. Phys. Lett.* **2007**, *91*, 202504. [CrossRef]
47. Chu, Y.H.; Martin, L.W.; Holcomb, M.B.; Gajek, M.; Han, S.J.; He, Q.; Balke, N.; Yang, C.H.; Lee, D.; Hu, W.; et al. Electric-field control of local ferromagnetism using a magnetoelectric multiferroic. *Nat. Mater.* **2008**, *7*, 478–482. [CrossRef] [PubMed]
48. Yadav, A.K.; Nelson, C.T.; Hsu, S.L.; Hong, Z.; Clarkson, J.D.; Schlepüetz, C.M.; Damodaran, A.R.; Shafer, P.; Arenholz, E.; Dedon, L.R.; et al. Observation of polar vortices in oxide superlattices. *Nature* **2016**, *530*, 198–201. [CrossRef] [PubMed]
49. Shafer, P.; García-Fernández, P.; Aguado-Puente, P.; Damodaran, A.R.; Yadav, A.K.; Nelson, C.T.; Hsu, S.L.; Wojdel, J.C.; Íñiguez, J.; Martin, L.W.; et al. Emergent chirality in the electric polarization texture of titanate superlattices. *Proc. Natl. Acad. Sci. USA* **2018**, *115*, 915–920. [CrossRef] [PubMed]
50. Khan, A.I.; Chatterjee, K.; Wang, B.; Drapcho, S.; You, L.; Serrao, C.; Bakaul, S.R.; Ramesh, R.; Salahuddin, S. Negative capacitance in a ferroelectric capacitor. *Nat. Mater.* **2015**, *14*, 182–186. [CrossRef] [PubMed]
51. Zubko, P.; Wojdeł, J.C.; Hadjimichael, M.; Fernandez-Pena, S.; Sené, A.; Luk'yanchuk, I.; Triscone, J.M.; Íñiguez, J. Negative capacitance in multi-domain ferroelectric superlattices. *Nature* **2016**, *534*, 524–528. [CrossRef] [PubMed]
52. Chu, Y.H.; Cruz, M.P.; Yang, C.H.; Martin, L.W.; Yang, P.L.; Zhang, J.X.; Lee, K.; Yu, P.; Chen, L.Q.; Ramesh, R. Domain Control in Multiferroic $BiFeO_3$ through Substrate Vicinality. *Adv. Mater.* **2007**, *19*, 2662–2666. [CrossRef]
53. Chu, Y.H.; He, Q.; Yang, C.H.; Yu, P.; Martin, L.W.; Shafer, P.; Ramesh, R. Nanoscale control of domain architectures in $BiFeO_3$ thin films. *Nano Lett.* **2009**, *9*, 1726–1730. [CrossRef] [PubMed]
54. Chen, Z.H.; Damodaran, A.R.; Xu, R.; Lee, S.; Martin, L.W. Effect of "symmetry mismatch" on the domain structure of rhombohedral $BiFeO_3$ thin films. *Appl. Phys. Lett.* **2014**, *104*, 182908. [CrossRef]
55. Haeni, J.H.; Irvin, P.; Chang, W.; Uecker, R.; Reiche, P.; Li, Y.L.; Choudhury, S.; Tian, W.; Hawley, M.E.; Craigo, B.; et al. Room-temperature ferroelectricity in strained $SrTiO_3$. *Nature* **2004**, *430*, 758–761. [CrossRef] [PubMed]
56. Vasudevarao, A.; Kumar, A.; Tian, L.; Haeni, J.H.; Li, Y.L.; Eklund, C.J.; Jia, Q.X.; Uecker, R.; Reiche, P.; Rabe, K.M.; et al. Multiferroic Domain Dynamics in Strained Strontium Titanate. *Phys. Rev. Lett.* **2006**, *97*, 257602. [CrossRef] [PubMed]
57. Fennie, C.J.; Rabe, K.M. Magnetic and Electric Phase Control in Epitaxial $EuTiO_3$ from First Principles. *Phys. Rev. Lett.* **2006**, *97*, 267602. [CrossRef] [PubMed]
58. Lee, J.H.; Fang, L.; Vlahos, E.; Ke, X.; Jung, Y.W.; Kourkoutis, L.F.; Kim, J.W.; Ryan, P.J.; Heeg, T.; Roeckerath, M.; et al. A strong ferroelectric ferromagnet created by means of spin-lattice coupling. *Nature* **2010**, *466*, 954–958. [CrossRef] [PubMed]
59. Highland, M.J.; Fister, T.T.; Fong, D.D.; Fuoss, P.H.; Thompson, C.; Eastman, J.A.; Streiffer, S.K.; Stephenson, G.B. Equilibrium Polarization of Ultrathin $PbTiO_3$ with Surface Compensation Controlled by Oxygen Partial Pressure. *Phys. Rev. Lett.* **2011**, *107*, 187602. [CrossRef] [PubMed]
60. Yu, P.; Luo, W.; Yi, D.; Zhang, J.X.; Rossell, M.D.; Yang, C.H.; You, L.; Singh-Bhalla, G.; Yang, S.Y.; He, Q.; et al. Interface control of bulk ferroelectric polarization. *Proc. Natl. Acad. Sci. USA* **2012**, *109*, 9710–9715. [CrossRef] [PubMed]
61. De Luca, G.; Strkalj, N.; Manz, S.; Bouillet, C.; Fiebig, M.; Trassin, M. Nanoscale design of polarization in ultrathin ferroelectric heterostructures. *Nat. Commun.* **2017**, *8*, 1419. [CrossRef] [PubMed]
62. Chen, D.; Chen, Z.; He, Q.; Clarkson, J.D.; Serrao, C.R.; Yadav, A.K.; Nowakowski, M.E.; Fan, Z.; You, L.; Gao, X.; et al. Interface Engineering of Domain Structures in $BiFeO_3$ Thin Films. *Nano Lett.* **2017**, *17*, 486–493. [CrossRef] [PubMed]
63. Lichtensteiger, C.; Weymann, C.; Fernandez-Pena, S.; Paruch, P.; Triscone, J.M. Built-in voltage in thin ferroelectric PbTiO3 films: The effect of electrostatic boundary conditions. *New J. Phys.* **2016**, *18*, 043030. [CrossRef]

64. Lichtensteiger, C.; Fernandez-Pena, S.; Weymann, C.; Zubko, P.; Triscone, J.M. Tuning of the Depolarization Field and Nanodomain Structure in Ferroelectric Thin Films. *Nano Lett.* **2014**, *14*, 4205–4211. [CrossRef] [PubMed]
65. Tang, Y.L.; Zhu, Y.L.; Ma, X.L.; Borisevich, A.Y.; Morozovska, A.N.; Eliseev, E.A.; Wang, W.Y.; Wang, Y.J.; Xu, Y.B.; Zhang, Z.D.; et al. Observation of a periodic array of flux-closure quadrants in strained ferroelectric $PbTiO_3$ films. *Science* **2015**, *348*, 547–551. [CrossRef] [PubMed]
66. Gruverman, A.; Wu, D.; Scott, J.F. Piezoresponse Force Microscopy Studies of Switching Behavior of Ferroelectric Capacitors on a 100-ns. *Phys. Rev. Lett.* **2008**, *100*, 097601. [CrossRef] [PubMed]
67. Jo, J.Y.; Yang, S.M.; Kim, T.H.; Lee, H.N.; Yoon, J.-G.; Park, S.; Jo, Y.; Jung, M.H.; Noh, T.W. Nonlinear Dynamics of Domain-Wall Propagation in Epitaxial Ferroelectric Thin Films. *Phys. Rev. Lett.* **2009**, *102*, 045701. [CrossRef] [PubMed]
68. Lefevre, C.; Demchenko, A.; Bouillet, C.; Luysberg, M.; Devaux, X.; Roulland, F.; Versini, G.; Barre, S.; Wakabayashi, Y.; Boudet, N.; et al. Nondestructive Method for the Determination of the Electric Polarization Orientation in Thin Films: Illustration on Gallium Ferrite Thin Films. *Small Methods* **2017**, *1*, 1700234. [CrossRef]
69. Lahtinen, T.H.E.; Tuomi, J.E.; van Dijken, S. Pattern Transfer and Electric-Field-Induced Magnetic Domain Formation in Multiferroic Heterostructures. *Adv. Mater.* **2011**, *23*, 3187–3191. [CrossRef] [PubMed]
70. Franke, K.J.A.; López González, D.; Hämäläinen, S.J.; van Dijken, S. Size Dependence of Domain Pattern Transfer in Multiferroic Heterostructures. *Phys. Rev. Lett.* **2014**, *112*, 017201. [CrossRef] [PubMed]
71. Franke, K.J.A.; Van de Wiele, B.; Shirahata, Y.; Hämäläinen, S.J.; Taniyama, T.; van Dijken, S. Reversible Electric-Field-Driven Magnetic Domain-Wall Motion. *Phys. Rev. X* **2015**, *5*, 011010. [CrossRef]
72. Trassin, M.; Clarkson, J.D.; Bowden, S.R.; Liu, J.; Heron, J.T.; Paull, R.J.; Arenholz, E.; Pierce, D.T.; Unguris, J. Interfacial coupling in multiferroic/ferromagnet heterostructures. *Phys. Rev. B* **2013**, *87*, 134426. [CrossRef]
73. Unguris, J.; Bowden, S.R.; Pierce, D.T.; Trassin, M.; Ramesh, R.; Cheong, S.W.; Fackler, S.; Takeuchi, I. Simultaneous imaging of the ferromagnetic and ferroelectric structure in multiferroic heterostructures. *APL Mater.* **2014**, *2*, 076109. [CrossRef]
74. You, L.; Wang, B.; Zou, X.; Lim, Z.S.; Zhou, Y.; Ding, H.; Chen, L.; Wang, J. Origin of the uniaxial magnetic anisotropy in $La_{0.7}Sr_{0.3}MnO_3$ on stripe-domain $BiFeO_3$. *Phys. Rev. B* **2013**, *88*, 184426. [CrossRef]
75. Maze, J.R.; Stanwix, P.L.; Hodges, J.S.; Hong, S.; Taylor, J.M.; Cappellaro, P.; Jiang, L.; Dutt, M.V.G.; Togan, E.; Zibrov, A.S.; et al. Nanoscale magnetic sensing with an individual electronic spin in diamond. *Nature* **2008**, *455*, 644–647. [CrossRef] [PubMed]
76. Balasubramanian, G.; Chan, I.Y.; Kolesov, R.; Al-Hmoud, M.; Tisler, J.; Shin, C.; Kim, C.; Wojcik, A.; Hemmer, P.R.; Krueger, A.; et al. Nanoscale imaging magnetometry with diamond spins under ambient conditions. *Nature* **2008**, *455*, 648–651. [CrossRef] [PubMed]
77. Rondin, L.; Tetienne, J.P.; Hingant, T.; Roch, J.F.; Maletinsky, P.; Jacques, V. Magnetometry with nitrogen-vacancy defects in diamond. *Rep. Prog. Phys.* **2014**, *77*, 056503. [CrossRef] [PubMed]
78. Gross, I.; Akhtar, W.; Garcia, V.; Martínez, L.J.; Chouaieb, S.; Garcia, K.; Carrétéro, C.; Barthélémy, A.; Appel, P.; Maletinsky, P.; et al. Real-space imaging of non-collinear antiferromagnetic order with a single-spin magnetometer. *Nature* **2017**, *549*, 252–256. [CrossRef] [PubMed]
79. Kaneshiro, J.; Uesu, Y.; Fukui, T. Visibility of inverted domain structures using the second harmonic generation microscope: Comparison of interference and non-interference cases. *J. Opt. Soc. Am. B* **2010**, *27*, 888–894. [CrossRef]
80. Lüpke, G. Characterization of semiconductor interfaces by second-harmonic generation. *Surf. Sci. Rep.* **1999**, *35*, 75–161. [CrossRef]
81. Fröhlich, D.; Kiefer, T.; Leute, S.; Lottermoser, T. Nonlinear spectroscopy of antiferromagnetics. *Appl. Phys. B* **1999**, *68*, 465–471. [CrossRef]
82. Rubano, A.; Fiebig, M.; Paparo, D.; Marino, A.; Maccariello, D.; di Uccio, U.S.; Granozio, F.M.; Marrucci, L.; Richter, C.; Paetel, S.; et al. Spectral and spatial distribution of polarization at the $LaAlO_3/SrTiO_3$ interface. *Phys. Rev. B* **2011**, *83*, 155405. [CrossRef]
83. Lummen, T.T.A.; Gu, Y.; Wang, J.; Lei, S.; Xue, F.; Kumar, A.; Barnes, A.T.; Barnes, E.; Denev, S.; Belianinov, A.; et al. Thermotropic phase boundaries in classic ferroelectrics. *Nat. Commun.* **2014**, *5*, 3172. [CrossRef] [PubMed]

84. Fiebig, M.; Fröhlich, D.; Sluyterman v. L., G.; Pisarev, R.V. Domain topography of antiferromagnetic Cr_2O_3 by second-harmonic generation. *Appl. Phys. Lett.* **1995**, *66*, 2906–2908. [CrossRef]
85. Schoenherr, P.; Giraldo, L.M.; Lilienblum, M.; Trassin, M.; Meier, D.; Fiebig, M. Magnetoelectric Force Microscopy on Antiferromagnetic 180° Domains in Cr_2O_3. *Materials* **2017**, *10*, 1051. [CrossRef] [PubMed]
86. Matsubara, M.; Manz, S.; Mochizuki, M.; Kubacka, T.; Iyama, A.; Aliouane, N.; Kimura, T.; Johnson, S.L.; Meier, D.; Fiebig, M. Magnetoelectric domain control in multiferroic $TbMnO_3$. *Science* **2015**, *348*, 1112–1115. [CrossRef] [PubMed]
87. Manz, S.; Matsubara, M.; Lottermoser, T.; Büchi, J.; Iyama, A.; Kimura, T.; Meier, D.; Fiebig, M. Reversible Optical Switching of Antiferromagnetism in $TbMnO_3$. *Nat. Photonics* **2016**, *10*, 653–656. [CrossRef]
88. Trassin, M.; De Luca, G.; Manz, S.; Fiebig, M. Probing Ferroelectric Domain Engineering in $BiFeO_3$ Thin Films by Second Harmonic Generation. *Adv. Mater.* **2015**, *27*, 4871–4876. [CrossRef] [PubMed]
89. Dick, B.; Gierulski, A.; Marowsky, G.; Reider, G.A. Determination of the nonlinear optical susceptibility $\chi^{(2)}$ of surface layers by sum and difference frequency generation in reflection and transmission. *Appl. Phys. B* **1985**, *38*, 107–116. [CrossRef]
90. Guyot-Sionnest, P.; Chen, W.; Shen, Y.R. General considerations on optical second-harmonic generation from surfaces and interfaces. *Phys. Rev. B* **1986**, *33*, 8254–8263. [CrossRef]
91. Nagarajan, V.; Jenkins, I.G.; Alpay, S.P.; Li, H.; Aggarwal, S.; Salamanca-Riba, L.; Roytburd, A.L.; Ramesh, R. Thickness dependence of structural and electrical properties in epitaxial lead zirconate titanate films. *J. Appl. Phys.* **1999**, *86*, 595–602. [CrossRef]
92. Nesterov, O.; Matzen, S.; Magen, C.; Vlooswijk, A.H.G.; Catalan, G.; Noheda, B. Thickness scaling of ferroelastic domains in $PbTiO_3$ films on $DyScO_3$. *Appl. Phys. Lett.* **2013**, *103*, 142901. [CrossRef]
93. Feigl, L.; Yudin, P.; Stolichnov, I.; Sluka, T.; Shapovalov, K.; Mtebwa, M.; Sandu, C.S.; Wei, X.K.; Tagantsev, A.K.; Setter, N. Controlled stripes of ultrafine ferroelectric domains. *Nat. Commun.* **2014**, *5*, 4677. [CrossRef] [PubMed]
94. Su, D.; Meng, Q.; Vaz, C.A.F.; Han, M.G.; Segal, Y.; Walker, F.J.; Sawicki, M.; Broadbridge, C.; Ahn, C.H. Origin of 90° domain wall pinning in $Pb(Zr_{0.2}Ti_{0.8})O_3$ heteroepitaxial thin films. *Appl. Phys. Lett.* **2011**, *99*, 102902. [CrossRef]
95. Gao, P.; Britson, J.; Jokisaari, J.R.; Nelson, C.T.; Baek, S.H.; Wang, Y.; Eom, C.B.; Chen, L.Q.; Pan, X. Atomic-scale mechanisms of ferroelastic domain-wall-mediated ferroelectric switching. *Nat. Commun.* **2013**, *4*, 2791. [CrossRef]
96. Gao, P.; Britson, J.; Nelson, C.T.; Jokisaari, J.R.; Duan, C.; Trassin, M.; Baek, S.H.; Guo, H.; Li, L.; Wang, Y.; et al. Ferroelastic domain switching dynamics under electrical and mechanical excitations. *Nat. Commun.* **2014**, *5*, 3801. [CrossRef] [PubMed]
97. De Luca, G.; Rossell, M.D.; Schaab, J.; Viart, N.; Fiebig, M.; Trassin, M. Domain wall architecture in tetragonal ferroelectric thin films. *Adv. Mater.* **2017**, *29*, 1605145. [CrossRef] [PubMed]
98. Zeches, R.J.; Rossell, M.D.; Zhang, J.X.; Hatt, A.J.; He, Q.; Yang, C.H.; Kumar, A.; Wang, C.H.; Melville, A.; Adamo, C.; et al. A Strain-Driven Morphotropic Phase Boundary in $BiFeO_3$. *Science* **2009**, *326*, 977–980. [CrossRef] [PubMed]
99. Sando, D.; Barthélémy, A.; Bibes, M. $BiFeO_3$ epitaxial thin films and devices: Past, present and future. *J. Phys. Condens. Matter* **2014**, *26*, 47320. [CrossRef] [PubMed]
100. Beekman, C.; Siemons, W.; Ward, T.Z.; Chi, M.; Howe, J.; Biegalski, M.D.; Balke, N.; Maksymovych, P.; Farrar, A.K.; Romero, J.B.; et al. Phase Transitions, Phase Coexistence, and Piezoelectric Switching Behavior in Highly Strained $BiFeO_3$ Films. *Adv. Mater.* **2013**, *25*, 5561–5567. [CrossRef] [PubMed]
101. Kumar, A.; Denev, S.; Zeches, R.J.; Vlahos, E.; Podraza, N.J.; Melville, A.; Schlom, D.G.; Ramesh, R.; Gopalan, V. Probing mixed tetragonal/rhombohedral-like monoclinic phases in strained bismuth ferrite films by optical second harmonic generation. *Appl. Phys. Lett.* **2010**, *97*, 112903. [CrossRef]
102. Haislmaier, R.C.; Podraza, N.J.; Denev, S.; Melville, A.; Schlom, D.G.; Gopalan, V. Large nonlinear optical coefficients in pseudo-tetragonal $BiFeO_3$ thin films. *Appl. Phys. Lett.* **2013**, *103*, 031906. [CrossRef]
103. Kumar, A.; Rai, R.C.; Podraza, N.J.; Denev, S.; Ramirez, M.; Chu, Y.H.; Martin, L.W.; Ihlefeld, J.; Heeg, T.; Schubert, J.; et al. Linear and nonlinear optical properties of $BiFeO_3$. *Appl. Phys. Lett.* **2008**, *92*, 121915. [CrossRef]
104. Zhang, J.; Ke, X.; Gou, G.; Seidel, J.; Xiang, B.; Yu, P.; Liang, W.I.; Minor, A.M.; Chu, Y.H.; van Tendeloo, G.; et al. A nanoscale shape memory oxide. *Nat. Commun.* **2013**, *4*, 2768. [CrossRef] [PubMed]

105. Seidel, J.; Trassin, M.; Zhang, Y.; Maksymovych, P.; Uhlig, T.; Milde, P.; Köhler, D.; Baddorf, A.P.; Kalinin, S.V.; Eng, L.M.; et al. Electronic Properties of Isosymmetric Phase Boundaries in Highly Strained Ca-Doped $BiFeO_3$. *Adv. Mater.* **2014**, *26*, 4376–4380. [CrossRef] [PubMed]
106. Roh, C.J.; Hamh, S.Y.; Woo, C.S.; Kim, K.E.; Yang, C.H.; Lee, J.S. Ferroelectric domain states of a tetragonal $BiFeO_3$ thin film investigated by second harmonic generation microscopy. *Nanoscale Res. Lett.* **2017**, *12*, 353. [CrossRef] [PubMed]
107. Wang, J.S.; Jin, K.J.; Guo, H.Z.; Gu, J.X.; Wan, Q.; He, X.; Li, X.L.; Xu, X.L.; Yang, G.Z. Evolution of structural distortion in $BiFeO_3$ thin films probed by second-harmonic generation. *Sci. Rep.* **2016**, *6*, 38268. [CrossRef] [PubMed]
108. Zhao, T.; Scholl, A.; Zavaliche, F.; Lee, K.; Barry, M.; Doran, A.; Cruz, M.P.; Chu, Y.H.; Ederer, C.; Spaldin, N.A.; et al. Electrical control of antiferromagnetic domains in multiferroic $BiFeO_3$ films at room temperature. *Nat. Mater.* **2006**, *5*, 823–829. [CrossRef] [PubMed]
109. Holcomb, M.B.; Martin, L.W.; Scholl, A.; He, Q.; Yu, P.; Yang, C.H.; Yang, S.Y.; Glans, P.A.; Valvidares, M.; Huijben, M.; et al. Probing the evolution of antiferromagnetism in multiferroics. *Phys. Rev. B* **2010**, *81*, 134406. [CrossRef]
110. Zhou, Z.; Trassin, M.; Gao, Y.; Gao, Y.; Qiu, D.; Ashraf, K.; Nan, T.; Yang, X.; Bowden, S.R.; Pierce, D.T.; et al. Probing electric field control of magnetism using ferromagnetic resonance. *Nat. Commun.* **2015**, *6*, 6082. [CrossRef] [PubMed]
111. Balke, N.; Choudhury, S.; Jesse, S.; Huijben, M.; Chu, Y.H.; Baddorf, A.P.; Chen, L.Q.; Ramesh, R.; Kalinin, S.V. Deterministic control of ferroelastic switching in multiferroic materials. *Nat. Nanotechnol.* **2009**, *4*, 868–875. [CrossRef] [PubMed]
112. Crassous, A.; Sluka, T.; Tagantsev, A.K.; Setter, N. Polarization charge as a reconfigurable quasi-dopant in ferroelectric thin films. *Nat. Nanotechnol.* **2015**, *10*, 614–618. [CrossRef] [PubMed]
113. Chauleau, J.Y.; Haltz, E.; Carrétéro, C.; Fusil, S.; Viret, M. Multi-stimuli manipulation of antiferromagnetic domains assessed by second-harmonic imaging. *Nat. Mater.* **2017**, *16*, 803–807. [CrossRef] [PubMed]
114. Meier, D. Functional domain walls in multiferroics. *J. Phys. Condens. Matter* **2015**, *27*, 463003. [CrossRef] [PubMed]
115. Seidel, J.; Martin, L.W.; He, Q.; Zhan, Q.; Chu, Y.H.; Rother, A.; Hawkridge, M.E.; Maksymovych, P.; Yu, P.; Gajek, M.; et al. Conduction at domain walls in oxide multiferroics. *Nat. Mater.* **2009**, *8*, 229–234. [CrossRef] [PubMed]
116. Mundy, J.A.; Schaab, J.; Kumagai, Y.; Cano, A.; Stengel, M.; Krug, I.P.; Gottlob, D.M.; Doğanay, H.; Holtz, M.E.; Held, R.; et al. Functional electronic inversion layers at ferroelectric domain walls. *Nat. Mater.* **2017**, *16*, 622–627. [CrossRef] [PubMed]
117. Sharma, P.; Zhang, Q.; Sando, D.; Lei, C.H.; Liu, Y.; Li, J.; Nagarajan, V.; Seidel, J. Nonvolatile ferroelectric domain wall memory. *Sci. Adv.* **2017**, *3*, e1700512. [CrossRef] [PubMed]
118. Holtz, M.E.; Shapovalov, K.; Mundy, J.A.; Chang, C.S.; Yan, Z.; Bourret, E.; Muller, D.A.; Meier, D.; Cano, A. Topological Defects in Hexagonal Manganites: Inner Structure and Emergent Electrostatics. *Nano Lett.* **2017**, *17*, 5883–5890. [CrossRef] [PubMed]
119. Tolédano, P.; Guennou, M.; Kreisel, J. Order-parameter symmetries of domain walls in ferroelectrics and ferroelastics. *Phys. Rev. B* **2014**, *89*, 134104. [CrossRef]
120. Becher, C.; Trassin, M.; Lilienblum, M.; Nelson, C.T.; Suresha, S.J.; Yi, D.; Yu, P.; Ramesh, R.; Fiebig, M.; Meier, D. Functional ferroic heterostructures with tunable integral symmetry. *Nat. Commun.* **2014**, *5*, 4295. [CrossRef] [PubMed]
121. Dennis, C.L.; Borges, R.P.; Buda, L.D.; Ebels, U.; Gregg, J.F.; Hehn, M.; Jouguelet, E.; Ounadjela, K.; Petej, I.; Prejbeanu, I.L.; et al. The defining length scales of mesomagnetism: A Review. *J. Phys. Condens. Matter* **2002**, *14*, R1175–R1262. [CrossRef]
122. Cherifi-Hertel, S.; Bulou, H.; Hertel, R.; Taupier, G.; Dorkenoo, K.D.; Andreas, C.; Guyonnet, J.; Gaponenko, I.; Gallo, K.; Paruch, P. Non-Ising and chiral ferroelectric domain walls revealed by nonlinear optical microscopy. *Nat. Commun.* **2017**, *8*, 15768. [CrossRef] [PubMed]
123. Lee, D.; Behera, R.K.; Wu, P.; Xu, H.; Sinnott, S.B.; Phillpot, S.R.; Chen, L.Q.; Gopalan, V. Mixed Bloch-Néel-Ising character of 180° ferroelectric domain walls. *Phys. Rev. B* **2009**, *80*, 060102. [CrossRef]

124. Flörsheimer, M.; Paschotta, R.; Kubitscheck, U.; Brillert, C.; Hofmann, D.; Heuer, L.; Schreiber, G.; Verbeek, C.; Sohler, W.; Fuchs, H. Second-harmonic imaging of ferroelectric domains in $LiNbO_3$ with micron resolution in lateral and axial directions. *Appl. Phys. B* **1998**, *67*, 593–599. [CrossRef]
125. Sheng, Y.; Best, A.; Butt, H.J.; Krolikowski, W.; Arie, A.; Koynov, K. Three-dimensional ferroelectric domain visualization by Čerenkov-type second harmonic generation. *Opt. Express* **2010**, *18*, 16539. [CrossRef] [PubMed]
126. Kämpfe, T.; Reichenbach, P.; Schröder, M.; Haußmann, A.; Eng, L.M.; Woike, T.; Soergel, E. Optical three-dimensional profiling of charged domain walls in ferroelectrics by Cherenkov second-harmonic generation. *Phys. Rev. B* **2014**, *89*, 035314. [CrossRef]
127. Wehmeier, L.; Kämpfe, T.; Haußmann, A.; Eng, L.M. In Situ 3D Observation of the Domain Wall Dynamics in a Triglycine Sulfate Single Crystal upon Ferroelectric Phase Transition. *Phys. Status Solidi (RRL)-Rapid Res. Lett.* **2017**, *11*, 1700267. [CrossRef]
128. Prylepa, A.; Reitböck, C.; Cobet, M.; Jesacher, A.; Jin, X.; Adelung, R.; Schatzl-Linder, M.; Luckeneder, G.; Stellnberger, K.H.; Steck, T.; et al. Material characterisation with methods of nonlinear optics. *J. Phys. D Appl. Phys.* **2018**, *51*, 043001. [CrossRef]
129. Neacsu, C.C.; van Aken, B.B.; Fiebig, M.; Raschke, M.B. Second-harmonic near-field imaging of ferroelectric domain structure of $YMnO_3$. *Phys. Rev. B* **2009**, *79*, 100107. [CrossRef]
130. Park, K.D.; Khatib, O.; Kravtsov, V.; Clark, G.; Xu, X.; Raschke, M.B. Hybrid tip-enhanced nanospectroscopy and nanoimaging of monolayer WSe_2 with local strain control. *Nano Lett.* **2016**, *16*, 2621–2627. [CrossRef] [PubMed]
131. Damodaran, A.R.; Clarkson, J.D.; Hong, Z.; Liu, H.; Yadav, A.K.; Nelson, C.T.; Hsu, S.L.; McCarter, M.R.; Park, K.D.; Kravtsov, V.; et al. Phase coexistence and electric-field control of toroidal order in oxide superlattices. *Nat. Mater.* **2017**, *16*, 1003–1009. [CrossRef] [PubMed]
132. Atkin, J.M.; Berweger, S.; Jones, A.C.; Raschke, M.B. Nano-optical imaging and spectroscopy of order, phases, and domains in complex solids. *Adv. Phys.* **2012**, *61*, 745–842. [CrossRef]
133. Zhuravlev, M.Y.; Sabirianov, R.F.; Jaswal, S.S.; Tsymbal, E.Y. Giant Electroresistance in Ferroelectric Tunnel Junctions. *Phys. Rev. Lett.* **2005**, *94*, 246802. [CrossRef]
134. Sanchez-Santolino, G.; Tornos, J.; Hernandez-Martin, D.; Beltran, J.I.; Munuera, C.; Cabero, M.; Perez-Muñoz, A.; Ricote, J.; Mompean, F.; Garcia-Hernandez, M.; et al. Resonant electron tunnelling assisted by charged domain walls in multiferroic tunnel junctions. *Nat. Nanotechnol.* **2017**, *12*, 655–662. [CrossRef] [PubMed]
135. Khan, A.I.; Marti, X.; Serrao, C.; Ramesh, R.; Salahuddin, S. Voltage-Controlled Ferroelastic Switching in $Pb(Zr_{0.2}Ti_{0.8})O_3$ Thin Films. *Nano Lett.* **2015**, *15*, 2229–2234. [CrossRef] [PubMed]
136. Sinsheimer, J.; Callori, S.J.; Ziegler, B.; Bein, B.; Chinta, P.V.; Ashrafi, A.; Headrick, R.L.; Dawber, M. In-situ X-ray diffraction study of the growth of highly strained epitaxial $BaTiO_3$ thin films. *Appl. Phys. Lett.* **2013**, *103*, 242904. [CrossRef]
137. Rubano, A.; Günter, T.; Lilienblum, M.; Aruta, C.; Granozio, F.M.; di Uccio, U.S.; Marrucci, L.; Paparo, D.; Fiebig, M. Optical second harmonic imaging as a diagnostic tool for monitoring epitaxial oxide thin-film growth. *Appl. Surf. Sci.* **2015**, *327*, 413–417. [CrossRef]
138. Jin, Q.Y.; Regensburger, H.; Vollmer, R.; Kirschner, J. Periodic Oscillations of the Surface Magnetization during the Growth of Co Films on Cu(001). *Phys. Rev. Lett.* **1998**, *80*, 4056–4059. [CrossRef]
139. Yamada, C.; Kimura, T. In-situ surface second-harmonic generation study of epitaxial growth of GaAs. *Jpn. J. Appl. Phys.* **1995**, *34*, 1102–1105. [CrossRef]
140. Jähnke, V.; Conrad, U.; Güdde, J.; Matthias, E. SHG investigations of the magnetization of thin Ni and Co films on Cu(001). *Appl. Phys. B* **1999**, *68*, 485–489. [CrossRef]
141. Hu, X.F.; Xu, Z.; Lim, D.; Downer, M.C.; Parkinson, P.S.; Gong, B.; Hess, G.; Ekerdt, J.G. In situ optical second-harmonic-generation monitoring of disilane adsorption and hydrogen desorption during epitaxial growth on Si(001). *Appl. Phys. Lett.* **1997**, *71*, 1376–1378. [CrossRef]
142. Roy, K.; Bandyopadhyay, S.; Atulasimha, J. Switching dynamics of a magnetostrictive single-domain nanomagnet subjected to stress. *Phys. Rev. B* **2011**, *83*, 224412. [CrossRef]
143. Hoffmann, T.; Thielen, P.; Becker, P.; Bohatý, L.; Fiebig, M. Time-resolved imaging of magnetoelectric switching in multiferroic $MnWO_4$. *Phys. Rev. B* **2011**, *84*, 184404. [CrossRef]

144. Miron, I.M.; Gaudin, G.; Auffret, S.; Rodmacq, B.; Schuhl, A.; Pizzini, S.; Vogel, J.; Gambardella, P. Current-driven spin torque induced by the Rashba effect in a ferromagnetic metal layer. *Nat. Mater.* **2010**, *9*, 230–234. [CrossRef] [PubMed]
145. Garello, K.; Avci, C.O.; Miron, I.M.; Baumgartner, M.; Ghosh, A.; Auffret, S.; Boulle, O.; Gaudin, G.; Gambardella, P. Ultrafast magnetization switching by spin-orbit torques. *Appl. Phys. Lett.* **2014**, *105*, 212402. [CrossRef]
146. Fiebig, M.; Lottermoser, T.; Fröhlich, D.; Goltsev, A.V.; Pisarev, R.V. Observation of coupled magnetic and electric domains. *Nature* **2002**, *419*, 818–820. [CrossRef] [PubMed]

© 2018 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (http://creativecommons.org/licenses/by/4.0/).

DiVA – Digitala Vetenskapliga Arkivet http://umu.diva-portal.org

This is an author produced version of a paper published in **Space Science Reviews**

This paper has been peer-reviewed but does not include the final publisher proof-corrections or journal pagination.

Citation for the published paper:

Ingrid Mann, Asta Pellinen-Wannberg, Edmond Murad, Olga Popova, Nicole Meyer-Vernet, Marlene Rosenberg, Tadashi Mukai, Andrzej Czechowski, Sonoyo Mukai, Jana Safrankova, Zdenek Nemecek

Dusty plasma effects in near earth space and interplanetary medium

Space Science Reviews, 2011, Vol. 161, Issue 1-4: 1-47

10.1007/s11214-011-9762-3

Access to the published version may require subscription. Published with permission from:

Springer Science+Business Media B.V.

Space Science Reviews manuscript No.
(will be inserted by the editor)

Dusty plasma effects in near Earth space and interplanetary medium

Ingrid Mann · Asta Pellinen–Wannberg · Edmond Murad · Olga Popova · Nicole Meyer–Vernet · Marlene Rosenberg · Tadashi Mukai · Andrzej Czechowski · Sonoyo Mukai · Jana Safrankova · Zdenek Nemecek

Received: date / Accepted: date

Abstract We review dust and meteoroid fluxes and their dusty plasma effects in the interplanetary medium near Earth orbit and in the Earth's ionosphere. Aside from in-situ measurements from sounding rockets and spacecraft, experimental data cover radar and optical observations of meteors. Dust plasma interactions in the interplanetary medium are observed by the detection of charged dust particles, by the detection of dust that is accelerated in the solar wind and by the detection of ions and neutrals that are released from the dust. These interactions are not well understood and lack quantitative description. There is still a huge discrepancy in the estimates of meteoroid mass deposition into the atmosphere. The radar

Ingrid Mann
Belgian Institute for Space Aeronomy
Tel.: + 32 2 3730406
Fax: + 32 2 3748423
E-mail: ingrid.mann@aeronomie.be
3 Avenue Circulaire, 1180 Brussels, Belgium

Asta Pellinen–Wannberg
Umeå University and Swedish Institute of Space Physics, Kiruna, Sweden

Edmond Murad
Newton, USA

Olga Popova
Institute for Dynamics of Geospheres, Russian Academy of Science, Moscow, Russia

Nicole Meyer–Vernet
Observatoire de Paris, Meudon, France

Marlene Rosenberg
Dept. of Electrical & Computer Engineering, Univ. of California, San Diego, USA

Tadashi Mukai
Kobe University, Japan

Andrzej Czechowski
Polish Space Research Institute, Warsaw, Poland

Sonoyo Mukai
Kinki University, Japan

Jana Safrankova and Zdenek Nemecek
Charles University, Prague, Czech Republic

meteor observations are of particular interest for determining this number. Dust measurements from spacecraft require a better understanding of the dust impact ionization process, as well as of the dust charging processes. The latter are also important for further studying nano dust trajectories in the solar wind. Moreover understanding of the complex dependencies that cause the variation of nano dust fluxes is still a challenge.

Keywords cosmic dust · meteors · aerosols · interplanetary medium · solar wind · ionosphere · Earth atmosphere · dusty plasma

1 Introduction

Plasmas containing charged dust particles of sizes $<$ mm are loosely referred to as dusty plasmas. The field of dusty plasma studies has grown rapidly over the last 15 years, driven by applications to astrophysical, space and laboratory plasmas. Dust is ubiquitous in the cosmos, occurring in interstellar, circumstellar, interplanetary, circumplanetary, and cometary environments. Dust also plays an important role in atmospheric processes. It is studied in laboratory plasmas, including plasmas used for materials processing, fusion plasmas, and dedicated dusty plasma experiments. Many processes in astrophysics rely on dust-plasma interactions and they often play a notable role, as, for instance, in the evolution of the interstellar medium.

Although the parameters of dusty plasmas cover an extremely wide range, certain fundamental concepts are common and may have application to many of these environments. For example, dust in a plasma is generally charged due to various mechanisms such as the collection of plasma electrons and ions. As another example, there are various forces that can act on the charged dust component, which are not usually considered for the much lighter ions and electrons such as gravitational forces, drag due to relative motion between dust and plasma or gas, and radiation pressure forces. And as a further example, charged dust can affect the plasma collective behavior, modifying waves and instabilities and leading to new very low frequency waves and instabilities associated with the dust.

Most of the dust-plasma interactions are at present not directly observed in space and the progress in the theory of dusty plasmas stands in contrast to the small amount of available data on the microphysics of dust-plasma interactions. The recent discussions of neutral solar wind, of inner source pick-up ions and of interplanetary field enhancements (see Section 4) suggest dust plasma interactions as a cause of these phenomena, but fail to provide the complete explanation.

Though the dust particles in the interplanetary medium and in near Earth space are immersed in plasmas with quite different parameters, what they have in common is that the space measurements provide direct access to them. The dust and larger meteoroids that cross the Earth orbit generate the dust in the ionosphere. The dust in the ionosphere is detected from sounding rockets and observed in meteors[1]

In Section 2 we describe the dust and plasma environment in the interplanetary medium and near Earth space. We discuss the dust flux onto the Earth and give an overview of

[1] According to the convention of the International Astronomical Union a meteor is in particular, the light phenomenon, which results from the entry into (usually the Earth's) atmosphere of a solid particle from space; and more generally any physical object or phenomenon associated with such an event. Meteoroids are bodies (usually in the interplanetary medium) that are considerably smaller than an asteroid and considerably larger than an atom or molecule. Meteorites are those parts of meteoroids that reach the surface of the Earth without being completely vaporized and "meteoric" is the adjectival form pertaining to meteors and meteoroids. Interplanetary dust particles are micrometeoroids typically of sizes of 100 μm and smaller. (cf. Mann 2009).

the different dust - plasma interactions. In Section 3 we describe dust measurements in the ionosphere based on observations of meteors, the light and ionization that is caused by meteoroids entering the Earth' s ionosphere. We further review the detection of dust by rocket-borne experiments and Radar and we discuss the possible formation of dust-plasma instabilities. In Section 4 we discuss the dust in the interplanetary medium near Earth orbit and the observation of dust generated neutrals and ions in the solar wind as well as the acceleration of nano dust in the solar wind. The Section 5 is devoted to dust charging and laboratory experiments related to dust in space. The physical principles of dust instruments are discussed in and Section 6 . Section 7 is a summary. A summarizing paragraph precedes each section. Except otherwise stated, SI units are used throughout.

2 The dust and plasma environments

The interplanetary medium includes a highly ionized solar wind and slowly flowing neutral gas. It is characterized by high plasma temperatures and small dust densities. There is a continuous distribution of objects from nm to the km sizes. The ionosphere has a low degree of ionization and a high gas density. Estimates of the amount of meteoroid material that enters from the interplanetary medium vary by orders of magnitude, but agree that the peak of this mass flux is approximately in the mass range of objects observed as radar meteors. The meteoroid material that remains in the atmosphere forms observed heavy ion layers. A fraction of this meteoroid material condenses into meteoric smoke particles. These are of nanometric size and were recently observed for the first time.

We describe the interplanetary medium in section (2.1) and the upper Earth' s atmosphere in section (2.2). Section (2.3) discusses the different estimates of the amount of meteoroid material that enters the Earth atmosphere and the final section (2.4) the possibly forming dust - plasma interactions.

2.1 Interplanetary medium

The solar wind plasma is the major component of the interplanetary medium. Following Meyer-Vernet (2007) we summarize some of its major characteristics. The solar wind has speeds on the order of 300-800 km/s and an average plasma density of 5×10^6 m $^{-3}$ near 1 AU (e.g. Table 2.4_2). In the fast wind, which comes from the solar polar regions, the density is $\sim 2-3 \times 10^6 \mathrm{m}^{-3}$ and the speed $\sim 700-800$ km/s. The slow wind has the density $\sim 7 \times 10^6$ m^{-3} and speed $\sim 300-500$ km/s. During ejections of coronal material, densities can reach $\sim 10^8$ m^{-3}. Near the Sun, in the F–corona, the slow wind can have speed ~ 100 km/s and density $\sim 10^{11}$ m^{-3} at around 4 solar radii, while coronal mass ejections are associated with denser plasmas that are faster than typical solar wind speed (see St Cyr et al. 2000). The solar wind drags the solar magnetic field, so that the magnetic field in the slow solar wind has a radial component of the order of 3 nT at 1 AU, whereas the solar rotation produces a spiral shape with an azimuthal magnetic field of roughly 3 nT at 1 AU in the ecliptic.

The vast majority of ions in the interplanetary medium are protons and α–particles, followed by highly ionized heavy ions. The measurements of particle fluences from roughly 0.3 keV/nucleon to 300 MeV/nucleon in the interplanetary medium were recently summarized by Mewaldt et al. (2007) and this gives the following picture: The solar wind contributes the majority of particles in the energy range up to ~ 8 keV/nucleon. It varies on a typical

scale of a factor of ~ 2 in velocity, flux, and composition and on time scales that range from hours to days to years. Intermediate range of energies between $\sim$ 10 keV/nucleon and $\sim$ 5 MeV/nucleon is covered by different types of particles and the fluences are more time-variable and difficult to predict. Solar energetic particles make the largest contribution at energies ~ 5 to ~ 50 MeV/nucleon, and galactic cosmic rays at even larger energies. The galactic and anomalous cosmic rays are rather stable in composition, while their energy spectra vary.

Asteroids and comets generate the majority of dust and meteoroids in the interplanetary medium and smaller dust particles subsequently form by collisional fragmentation. The interstellar dust provides a further component (see e.g. Mann et al. 2004). The dust velocities relative to their parent bodies are typically comparatively small and dust particles initially move with the velocities of the (Keplerian) orbits of the parent-bodies. The lifetime of dust roughly larger than μm is limited by mutual collisions and by the Poynting-Robertson effect (see 4.3) that results from radiation pressure force. Smaller particles are in majority ejected from the inner solar system.

2.2 The upper Earth' s atmosphere

The Earth is shielded from the solar wind by its magnetic field that gives rise to the magnetospheres which ranges roughly 10 to 12 earth radii (70000 km) from the center of the Earth in the sunward direction and more than 200 Earth radii in the direction away from the Sun. While it shields the Earth's atmosphere from the solar wind, meteoroids and dust can freely enter. Meteors are observed from about 70 km to about 400 km (see Section 3). This altitude range covers most of what is typically referred to as "ionosphere" and comprises the D, E and F layers. Transitions between those layers are variable. The layers were initially distinguished by their distinct reflections of radio waves and their composition is influenced by absorption of solar radiation. The D region extends below 90 km, the E region between 90 and 150 km and the F region between 150 and 500 km (following Kelley 1989).

In terms of temperature the region of interest lies mainly within the thermosphere. This is the region above the mesopause, which is defined as the altitude of minimum temperature. Temperatures, as well as the location of the mesopause are highly variable. LIDAR observations at two mid-latitude sites during night time, for instance, give mesopause altitudes of 101 km in winter and at 85 km during summer, while in spring and autumn the altitude oscillates between these states (Plane et al. 1999). The plasma is generally cold, with temperatures ranging from about a few hundred K at low altitudes (or even colder in the summer mesopause) to about 2000 K at higher altitudes in daytime, and about 1000 K in nighttime. Electron temperature is typically larger than ion and neutral temperatures.

The altitude profiles of the dominant neutral atmospheric constituents N_2, O_2, O as well as the dominant ionized constituents NO^+, O_2^+, O^+ and their variations due to the solar cycle, time of day, latitude, season and geomagnetic activity is quite well known above about 80 km. The densities are highly variable from day to night and from year to year, most notably with solar cycle. Fig. 1 shows the profiles of these neutral and ionized atmospheric constituents in solar maximum and minimum conditions at night. The plasmas are weakly ionized (see Table 2.4_2). Note, that the lower part of the region that we consider is also the region where ionized species are molecular, primarily NO^+; at altitudes lower than 80 km water hydrates such as $H^+(H_2O)_{n>2}$ assume increasing importance and replace other diatomic ions (Kull et al. 1997).

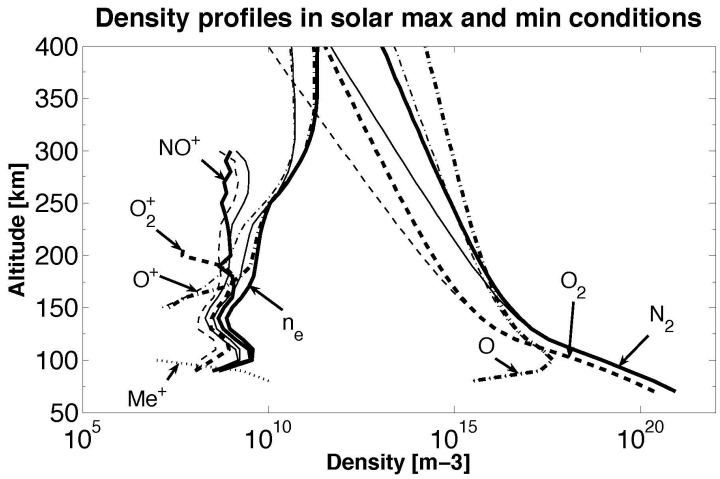


Fig. 1 Density profiles of the neutral atmosphere constituents N_2 (solid), O_2 (dashed) and O (dashed-dotted) under solar maximum (bold) and minimum conditions are located to the right in the density scale. Similarly the ionized components are shown as the total electron density profiles n_e (solid) in the middle and NO^+ (solid), O_2^+ (dashed) and O^+ (dash-dotted) to the left. The dotted profile represents ions of meteoric origin, Me^+. These profiles are generated from the IRI model (Bilitza and Reinisch 2008) for the ionosphere and MSIS model (Hedin 1991) for the atmosphere under solar maximum conditions in December 1990 and solar minimum conditions in December 1996 at location 60^0 N, 20^0 E at 22 UT. Densities are higher during daytime. The plasma (electron) density for the the same IRI model and day at 13.5 hours local time ranges from $7 \cdot 10^9 m^{-3}$ at 90 km to $10^{12} m^{-3}$ at 400 km . Neutral densities (from the MSIS model) range from about $5 \cdot 10^{18}$ m^{-3}at 90 km to $2 \cdot 10^{13}$ m^{-3} to $2 \cdot 10^{14}$ m^{-3} at 400 km (with the lower number for 1996). The meteoric ion profile is taken from Roble (1995).

2.3 Dust flux in the solar wind and into Earth atmosphere

The different estimates of meteoroid flux onto Earth vary by more than an order of magnitude, from 2 to $80 \cdot 10^7$ kg yr^{-1} over the whole planet (note though that observed meteor fluxes vary significantly with location, diurnal and annually, see e.g. Szasz et al. 2004). They are based on studies of dust and meteoroids outside of the atmosphere (Hughes 1978; Grün et al. 1985; Love and Brownlee 1993; and others), on atmospheric studies, as well as on studies of the deposition of extraterrestrial matter in glaciers and sediments (see Kane and Gardner 1993; Taylor et al. 1998; Mathews et al. 2001; Cziczo et al. 2001; Gabrielli et al. 2004; and others). Aside from the differences between the different observation methods, the uncertainties, as pointed out by Dyrud et al. (2008), arise from the characteristics of different radar meteor observation techniques.

The cumulative mass distribution of objects crossing the Earth orbit (see Fig. 2) increases towards smaller sizes roughly as a power law. For instance Ceplecha et al. (1998)derived a flux curve based on compiled data from meteor observations, dust in–situ measurements from spacecraft, analyses of lunar craters, as well as near–Earth asteroid observations.

Table 2.3_1 The mass range of dust and meteor observations in the ionosphere and near Earth space, the given mass ranges of different observations are discussed in the text.

Observation	Detected species	Object mass / size
UV photometers on rockets	scattered sunlight	size > 40 nm
Ionospheric dust detectors	dust surface charges	10^{-24}–10^{-23} kg
Plasma wave instruments	impact generated charges	$< 10^{-20}$–10^{-12} kg
Space dust detectors	impact generated charges	10^{-18}–10^{-12} kg
Head echo observations	free electrons	10^{-9}–10^{-6} kg
Radar meteor observations	free electrons	10^{-12}–10^{-5} kg
Optical meteor observations	meteoroid & atmospheric molecules, atoms & ions	10^{-8} kg – 10^{-1} kg
Fireballs observations	meteoroid & atmospheric molecules, atoms & ions	10^{-1} kg – 10^{3} kg

The derived average flux over 100 years of meteoroids with masses $<10^8$kg is $2.4 \cdot 10^7$ kg yr^{-1}, with masses $<10^{-2}$ kg is $4.2 \cdot 10^6$ kg yr^{-1} and with masses $<10^{-7}$ kg is $4 \cdot 10^6$ kg yr^{-1}. An analysis of impact craters collected during an exposure experiment on a near–Earth satellite Long Duration Exposure Facility in the 1980s leads to a larger flux in the size range 20 to 400 μm (approximate mass range 10^{-12}–10^{-7} kg) (Love and Brownlee 1993). Analysis of marine isotope records leads to a long term average accretion rate over the last 80 million years of $3.7 \cdot 10^7$ kg yr^{-1} (Peucker–Ehrenbrink 1996). We follow here the curve of Ceplecha et al., since it covers a broad size interval as well as observations of the past decades and assume that the flux amounts to 10 ton/day or more locally and temporarily.

Note that the mass distribution of objects smaller than 100 g peaks at roughly 10^{-8} kg (see Fig. 3 in Grün et al. 1985), implying that radar observations cover that interval in the dust and meteoroids population that causes most of the mass deposition into the atmosphere. Dust particles in the approximate size interval 1 to 100 μm are observed by the thermal emission brightness and scattered light brightness in the Zodiacal light. They are well determined near the ecliptic between 0.3 and 1.7 AU distance from the Sun; the Zodiacal light is stable in time (cf. Leinert et al. 1998). In-situ measurements from spacecraft suggest that the fluxes are stable in time for dust above 1/10 micrometer size (cf. Krüger et al. 2006). In contrast to that the flux of nano dust is highly time variable (see 4.3).

Dust composition is inferred from the composition of its parentbodies, from the composition of interplanetary dust collected in the Earth atmosphere, from the analysis of micrometeorites and meteorites collected on Earth, and from optical properties derived from Zodiacal light observations. Collected samples provide compositional information and show that most particles are heterogeneous in material composition and structure and that a significant fraction of dust is porous (see e.g. Rietmeijer 2002). For a rough estimate of the materials enter the Earth atmosphere, we assume that the majority of entering objects have cometary dust composition, similar to a local interstellar cloud dust model that was derived from astronomical observations and space measurements (Kimura et al. 2003). This model assumes that the elements are contained in the typical cosmic materials corundum (Al_2O_3), kamacite (FeNi), forsterite (Mg_2SiO_4), enstatite ($MgSiO_3$), troilite (FeS) and organic refractory compounds. Following the element abundances assumed in this model the mass percentages of major elements deposited into the ionosphere are: carbon 25%, nitrogen 4.5%, oxygen 38%, magnesium 7.1%, aluminium <1%, silicon 7.6 %, sulphur 3.2% and iron 13%. The amount of hydrogen and noble gases in cosmic dust and meteoroids is not well determined.

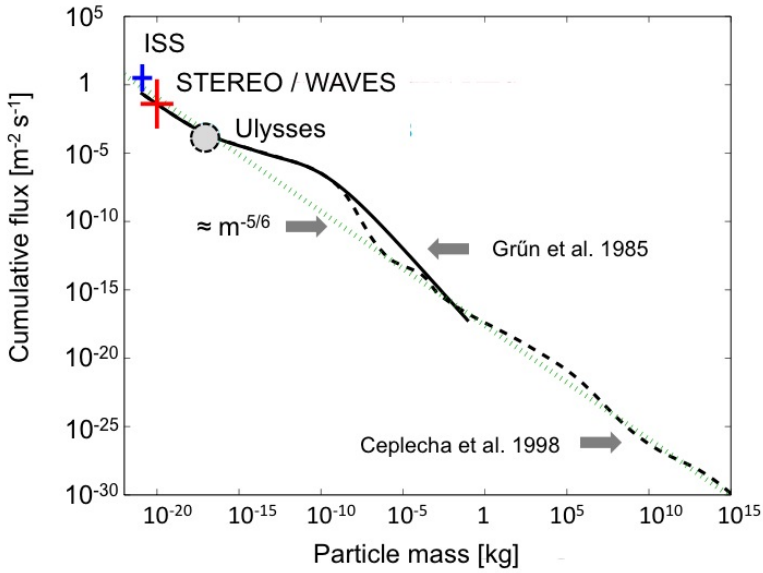


Fig. 2 The flux of dust with mass greater or equal the given mass near 1 AU. Shown are the interplanetary dust flux model that is derived primarily from dust in-situ measurements and analysis of lunar microcraters (solid line, (Grün et al. 1985)) and the meteor model that also includes the small solar system objects (dashed, (Ceplecha et al. 1998)). The dotted line describes a flux $\sim m^{-5/6}$ from collisional fragmentation equilibrium (Dohnanyi 1978). The flux values derived from analysis of impact craters on the International Space Station (ISS, Carpenter et al. 2005), from measurements of the Ulysses dust instrument (Wehry and Mann 1999) and and from STEREO plasma wave measurements during the year 2007 are shown. This figure is adapted from Meyer-Vernet et al. (2009b).

2.4 Dust plasma interactions

Dust can interact with a plasma in various ways. Dust can act as a sink of plasma particles by absorbing plasma species. Dust can also act as a source of electrons, via processes such as photoemission due to solar ultraviolet radiation (see section 5.1). The ablation of heated dust or meteors can release dust and atomic/molecular species into the plasma (see section 3.1). Impinging particles are decelerated and undergo charge exchange or may remove atoms or ions from the dust (i.e. sputtering) (see section 4.1).

Dust in a plasma and radiative environment is generally electrically charged owing to various processes, such as the collection of plasma particles, photoemission, secondary emission in the presence of high energy plasma particles, etc. (see section 5.1). When the spacing between dust particles is much larger than the Debye length, the charged dust particles are shielded out to such an extent that the electrostatic interaction between dust particles is negligible. On the other hand, when the spacing between dust particles is smaller than or comparable to the Debye length, the electrostatic interaction between dust particles is important. In this case the dust possibly gives rise to collective effects in the “dusty plasma”, see reviews by Mendis and Rosenberg 1994; Fortov et al. 2005; Shukla and Mamun 2002).

Both theoretical and laboratory experimental studies show that the presence of charged dust can modify known plasma waves and instabilities and give rise to new very low frequency waves and instabilities associated with the massive dust grains (for reviews see e.g.

Table 2.4_2 Parameters of dust and gas in the interplanetary medium near Earth orbit and in the ionosphere. Dust masses $m < 10^{-13}$ kg correspond to approximate sizes $< 2\mu$m, dust masses $m < 10^{-20}$ kg to sizes < 10 nm. Basic solar wind parameters are from Allen's Astrophysical Quantities (Allen 1976). The density of nano dust near 1 AU is estimated assuming the flux $\phi = 0.6m^{-2}s^{-1}$ and velocity 300 km/s for dust with mass 10^{-20} kg, the number densities of larger dust are from Mann et al. (2004). The ionospheric densities, temperatures and ionization fractions are discussed in Section (2.2). Numbers vary significantly and these are therefore rough estimates. The dust densities in the ionosphere are from the observations cited in this text (see Section 3.7 and Fig.5)

	Interplanetary Medium	Earth Ionosphere
	Medium	Ionosphere
M_{gas} / kg m^{-3}	$\approx 0.1 - 1 \times 10^{-19}$	$> 5 \times 10^{-16}$
M_{dust} / kg m^{-3}	2×10^{-18}	(not measured)
n_{dust} /m^{-3} ($m < 10^{-13}$ kg)	1×10^{-7}	(not measured)
n_{dust} /m^{-3} ($m < 10^{-20}$ kg)	$\approx 10^{-6}$	$(10^6 - 10^9)$
n_e /n_n	≈ 1	$\approx 10^{-10} - 10^{-3}$
n_e / m^{-3}	5×10^6	$10^9 - 10^{15}$
T_e / K	2×10^5	< 1500

Merlino et al, 1998; Verheest 2000, Shukla and Mamun, 2002; Fortov et al., 2005; Shukla and Eliasson, 2009).

For example, dust can affect standard waves and instabilities through effects on equilibrium charge neutrality, given by $n_e + \varepsilon Z_d n_d = n_i$, where Z_d is the dust charge state, $\varepsilon = 1(-1)$ for negatively (positively) charged dust, and n_d, n_e, n_i are the dust, electron and ion densities, respectively.

The phase speed of ion acoustic waves, which is $\propto (n_i/n_e)^{1/2}$, increases in the presence of negatively charged dust, so that the waves undergo less Landau damping and can be more easily excited (Shukla and Silin 1992; D'Angelo 1990; Rosenberg 1993). As another example, the motion of charged dust leads to a new very low frequency dust acoustic wave, where the dust provides the inertia and the electrons and ions provide the pressure to sustain the wave (Rao et al, 1990). Dust acoustic waves have phase speed less than the ion thermal speed and can be driven unstable by the streaming of ions (or electrons) with small speeds relative to the dust (Rosenberg, 1993). Although potentially interesting, the application of possible instabilities to near-Earth dusty plasmas is in a relatively early stage (see section 3.8).

3 Dust in the ionosphere

Entrance of large enough particles of extraterrestrial matter into the Earth' s atmosphere is associated with light emission and ionization phenomena giving rise to meteors. A region of ionized gas forms near the meteoroid and expands almost instantaneously. The head moves with the speed of the meteoroid. It is followed by the meteor trail (also called train), that extends behind the meteoroid body. At typical heights between 80 and 120 km, both head and trail, depending on meteoroid mass and entry conditions, can be observed optically as well as by radar backscattering. The relation between observed brightness and ablated meteoroid mass varies with the atmospheric conditions and with meteoroid parameters like altitude, entry speed, size and composition. Because of this variation physical parameters

like the bulk density can not be inferred completely from the observations. Meteoric smoke particles form from the ablated meteoroid material and are detected in-situ from sounding rockets. Charged dust particles are suspected to influence some meteor observations and be observed by radar scattered signals, but distinguishing between the heavy ion and charged dust signals is still a challenge. Observations of dusty plasma effects are still limited. Our discussion is primarily based on meteor phenomena and we refer the reader to detailed reviews of the particle interactions in the PMSE region (e.g. Cho and Röttger 1997, Rapp and Lübken 2004).

We start by describing the classical ablation process by evaporation (3.1). This explains most radar meteors (section 3.2), and optical meteors (see 3.3). High altitude meteors are discussed in section 3.4. They are associated to non-thermal ablation (3.5) especially due to sputtering. Fragmentation events, as described in section 3.6 limit the interpretation of meteor observations. The section 3.7 is devoted to atmospheric dust measurements and section 3.8 elaborates the possibility of dust-plasma instabilities forming in the Earth' s ionosphere.

3.1 Meteoroid ablation by evaporation

Meteoroids entering the Earth' s atmosphere lose mass mostly as a result of vapor production. The intensity of radiation I and the linear electron concentration α (i.e. number of free electrons produced by evaporated meteoroid atoms per meter), given in the following analytical expressions, can both be observed in the meteor phenomena:

$$I = -\tau\left(\frac{v^2}{2}\frac{dm}{dt} + mv\frac{dv}{dt}\right) \qquad \alpha = -\frac{\beta}{\mu v}\frac{dm}{dt} \tag{1}$$

where v, m are the meteoroid velocity and mass, τ the luminous efficiency, β the ionizing efficiency, and μ is the average mass of the ablated atoms (Ceplecha et al. 1998). The meteoroids size relation to the atmospheric free path at different altitudes corrected for meteoroid velocity defines the regime of interaction with ionosphere and the character of ablation (Bronshten 1983; Popova 2004). Large meteoroids lose most of their mass in the continuum flow regime whereas small meteoroids, discussed in this paper, interact with ionosphere mainly in the free molecular flow or transition flow regimes. (Note that, since the difference between thermal velocity and meteoroid velocity influences the physics of the interaction process, these boundaries are dependent both on size and velocity of meteoroid). The screening of meteoroid surface by evaporated material and subsequent decrease of the heat transfer coefficient must be taken into account in the transition flow condition modeling (Popova 2004 and references therein).

Momentum transfer in collisions between atmospheric gases and meteoroids leads to heating, followed by ablations and thermal radiation, the latter effects causing the meteoroid to cool. At a given altitude the behavior of a particle penetrating the ionosphere is determined by its size, velocity and material properties and this is shown in Fig. 3. Based on meteor physics equations analysis similar to that introduced the 1960s (see e.g. Popova 2004 and references therein), it shows that for impact speed of 40 km/s, stony objects ($\leq 10^{-6}$ m), decelerate before being substantially heated and the heating of objects with radiua R $\leq 10^{-4}$ m is limited by thermal radiation. Larger particles evaporate almost totally at altitudes between 100 and 80 km.

The recent analysis of micrometeoroids collected on the arctic surface snow point to mechanical destruction and weathering (Duprat et al. 2007). Rietmeijer (2002) points out that even certain classes of the interplanetary dust particles that are collected - almost intact

- in the Earth atmosphere show traces of flash heating to 300 - 1000 degree Celsius. Organic material may survive the entry process in the small decelerated particles which were not heated up to high temperatures (see e.g. Coulson 2002, Coulson and Wickramasinghe 2003). Besides, the more volatile metals may sublimate earlier during the meteoroid entry (differential ablation, e.g. McNeil et al., 1998). The heating of larger objects (R $>10^{-3}$ m) cannot be characterized by uniform temperature and the heated layer is about 10^{-4} – 10^{-3} m in thickness (Bronshten 1983; Ceplecha et al. 1998).

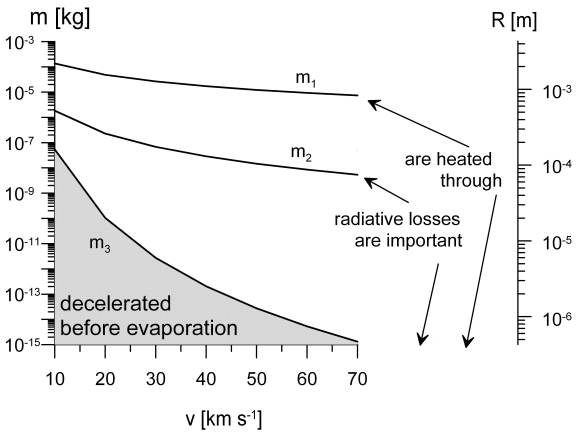


Fig. 3 Rough mass and size boundaries demonstrating importance of different processes.

The energy flux on the meteoroid increases as the body reaches denser regions of the atmosphere. Intensive evaporation starts when the temperature rise up to about 2000 – 2500 K, which occurs at about 110-130 km for porous bodies with R $\sim 10^{-3}$–10^{-1} m (Bronshten 1983; Ceplecha et al. 1998). Below this altitude the incoming energy contributes to ablation mainly. Above a certain size limit, which is dependent on material properties and velocity, the entering objects do not fully ablate (roughly objects larger 1-10 cm for low velocity entry). The remnant objects continue a dark flight and eventually reach the surface of the Earth as meteorites after undergoing a variety of heterogeneous reactions. The strongly temperature related meteoroid mass loss processes are generally called thermal ablation and other processes are usually excluded from the consideration. Aside from the thermal ablation, dedicated fragmentation models and the different conditions needed for high altitude meteors are discussed below.

3.2 Radar meteor observations

The meteor trail echoes are radio waves scattered by free electrons in the trail that extends for meters to kilometers behind the meteoroid. They last from tenths of seconds to minutes and dominate measurements by specular meteor radars (SMR) such as AMOR and CMOR (see Pellinen-Wannberg et al. 2008, for an overview of radar facilities).

According to their distinctly different shapes of the radar signal, trails are denoted as over–dense and under–dense trails. In the case of an under–dense meteor trail the plasma frequency is below the radar frequency. The scattered signals result from individual free

electrons within the trail and are reinforced due to constructive interference if the trail is oriented perpendicularly to the radar beam. Since the beam is wide, many trails fulfill the perpendicularity condition. When the plasma (electron) density is high, so that the plasma frequency exceeds the radar frequency, the radio waves can not penetrate the trail, which then forms a reflecting obstacle to the radar beam. This case is denoted as over–dense trail. The over-dense echo power is approximately constant until the expanding trail reaches the underdense condition, i.e. the column electron density along the beam is $<10^{14}$ m^{-1} (McKinley 1961) (the exact value depending on the radar frequency).

Theoretical predictions of the *initial radius* of evolving plasma column lead to differing results. Experimentally derived values range from several 10 cm to about 3 m depending on altitude and meteoroid velocity (Ceplecha et al. 1998). The initial radius (similarly as the mean free path of atmospheric species) increases with altitude. Hence, the phase difference between the scattered signals increases. As soon as the trail extends to more than half the radar wavelength constructive interference ceases, signals interfere destructively and the measured reflected power decreases exponentially. Hence the reflective properties of the meteor trail and the surrounding ionosphere for a given radio frequency/wavelength limit the height of SMR observations. This is called the radio meteor ceiling effect.

Head echoes originate from the plasma in the immediate surrounding of the meteoroid and detect the submillimeter-size meteoroid population in the 10^{-9} to 10^{-6} kg mass range. They are transient, typically lasting for less than tens of milliseconds, and characterised by high Doppler shifts. Only the most powerful SMRs are sensitive enough to observe such echoes and 700 head echoes were reported during 25 years of data on different showers (Jones and Webster 1991). Head echoes dominate measurements of High Power Large Aperture radars (HPLA), such as ALTAIR, Arecibo, and EISCAT (Pellinen-Wannberg et al. 2008), and they were firstly studied with EISCAT in 1990 (Pellinen–Wannberg and Wannberg 1994). The most powerful HPLA radars record 700 head echoes of sporadic meteors within a quarter of an hour (Close et al. 2000).

With head echo observations, parameters such as velocity and deceleration components along the radar beam, altitude of occurrence and rates can be obtained directly (Kero et al. 2008a). Tristatic measurements allow observing fragment formation (Kero et al. 2008b) and estimating the meteoroid mass and size with high accuracy (Kero et al. 2008c). These measurements also lead to a better derivation of the magnitude and orbit determinations for individual meteors (Szasz et al. 2007; Szasz et al. 2008). Janches et al. (2009) even suggest that head echo observations allow observing differential ablation. Observations of a single head echo at two or even three frequencies are rare, but allow closer studies of the head echo character (Wannberg et al. 1996; Close et al. 2002; Pellinen–Wannberg 2004).

3.3 Optical meteor observations

The most common optical meteor phenomena, their formation mechanisms and the main conditions of appearance are listed in Table 3.3_1 (from Mann 2009). The approximate mass ranges are for optical meteors $10^{-8} < m < 10^{-1}$ kg and for fireballs $10^{-1} < m < 10^{3}$ kg. The immediate surrounding of the meteoroid body displays the brightest part of the meteor, the head. Behind the head, different types of trails are observed and arise from different physical processes. The listed trail phenomena were discussed in detail by (Borovička 2006). Meteor spectra reveal mostly emission lines as well as some molecular bands; thermal continuum radiation is typically negligible and only dominant in the fireball spectra (e.g. Borovička 1999). Spectra change along meteor trajectory. The emission is due mostly to neutral atoms

Table 3.3_1 The major optical meteor phenomena. This table is adapted from Mann (2009), it basis on a detailed discussion of the trail phenomena by Borovička (2006).

Type	Mechanism(s)	Phenomenon	Main conditions of appearance
head	instantaneous line emission of meteoroid and atmospheric species surrounding the meteoroid	major meteor brightness that lasts for about a second	altitude between 120 and 80 km (20 km for fireballs)
flare	sudden fragmentation or change of physical parameters	sudden increase of meteor brightness by more than one magnitude	mostly for high velocity meteors
trail			
gaseous wake	line emission of rarefied non–equilibrium gas that follows the head	emission spectrum similar, but not identical to the head evolves 100s of meters to several kilometers behind the head	altitude beyond 55 km
particulate wake	individual meteor heads formed around fragments released from the meteoroid	emission spectrum similar to meteor head	below the altitude of fragmentation (mostly at low altitude)
green train	reactions among atmospheric species	OI 557.7 nm (forbidden line) emission occuring from 1 to 2 s after meteor disappearance	mostly for high velocity meteors; height profile different from that of the meteor
persistent train	atomic recombination and/or chemiluminescence of cooling rarefied gas	visible for tens of minutes after meteor disappearance; consist of afterglow phase, recombination phase and continuum	bright fireballs 75 to 100 km altitude
reflection train	Sunlight scattered at fragment dust particles generated by fireballs	bright spectral continuum and band emission of metal oxides that lasts for hours	form at any height; observed during daytime or twilight

and singly charged atomic ions (see Borovička 1993 for a compilation of the observed lines). The observed spectra are empirically described by two gas components with temperature $\sim$4000 K for the main component and $\sim$10 000 K for a second component. This model matches most observed spectra. In a typical application of this model, the emission lines of the main component are used to derive the chemical composition of the entering meteoroids (Borovička 1993). This empirical model, however, uses simplified description of the luminous volume and doesn't take into account possible deviation from thermodynamical equilibrium. Moreover the observed line emission partly originates from atmospheric, as opposed to meteoroid species (Campbell–Brown 2004; Hawkes 2002; Lindblad et al. 2003).

3.4 High altitude meteors

According to the scenario outlined above, meteor ablation starts below 130 km. Indeed, among more than 4500 photographic meteors recorded in the Meteor Data Center of the International Astronomical Union (Lindblad et al. 2003) only 0.1 % (five cases) had beginning

heights above 130 km and average beginning heights are below 100 km (Koten et al. 2006). Beginning heights increase with entry velocities. Indeed Campbell et al. (2000) derived a mean beginning height of 112 km for Leonids (v $\sim$ 72 km/s) with average mass of $1.4 \cdot 10^{-6}$ kg in 1998 (79 meteors in the sample). Brown et al. (2002) found that the ablation zone begun at about 123 km altitude for meteors with average mass of $\sim 10^{-6}$–10^{-7} kg in 1999 (232 Leonids).

Also, high altitude meteors, i.e. those observed beyond classical ablation heights, are often observed for meteoroids with high entry velocity. Koten et al. (2006) found about 5% high– altitude meteors, with beginning heights between roughly 130 and 150 km. among 164 observed shower meteors. But not a single Orionid was detected above 130 km despite their having a high entry velocity (v $\sim$ 66.5 km/s) The beginning height obviously varies with the spectral range and the sensitivity of the observation device. Simultaneous observations of two high-altitude Leonids in 1995 and 1996 showed beginning heights below 130 km for the photographic observations and around 160 km for intensified video observations (Fujiwara et al. 1998). Studying the high altitude spectra of 8 Leonids Spurný et al. (2004) find that the high altitude signal is mainly due to emission of the oxygen line at 777 nm together with a faint continuum that they suggest to be due to N_2 molecular bands.

Another Leonid observation from 2002 shows a trail starting at 145 km altitude (Pellinen–Wannberg et al. 2004). It was detected with the optical auroral imaging network ALIS (Brändström 2003) through two filters at 589.3 ± 10 nm in the range of a sodium emission line (NaI) and at 422.7 ± 14 nm in the range of calcium line (CaI). While the sodium trail started at 110 km, the signal through the calcium filter started at 145 altitude, was transient and very strong (Fig. 4). In addition to calcium this filter covered emission lines from iron and the hydrogen γ Balmer series. Pellinen–Wannberg et al. (2004) suggested that since calcium and iron lines were expected to be weaker, the emission originated from the γ Balmer series of a hydrogen compound induced by hypervelocity collisions.

Pellinen–Wannberg and Wannberg (1994) reported the detection of radar meteor trails in 1990 up 160 km altitude and later suggested that heavy meteoric ions modify the incoherent scatter spectrum from these trails (Pellinen–Wannberg and Wannberg 1996). During the 1998 Leonids Brosch et al. (2001) monitored echoes from meteors up to 400 km altitudes. They have interpreted the origin of these trails as sputtering of atmospheric particles on meteoroids at these altitudes.

Brosch et al. (2004) presented observations of simultaneous radar and optical Leonids and among that showed a radar event at very high altitude ($\sim$250 km), with a sequence of echoes indicating a sinuous path of the radar reflector. It was hypothesized that the radar scattering might occur from charged dust clouds, whose diffusion would be much smaller than ions owing to the large dust mass, and which could be carried by high–altitude winds. If dust is involved, one may surmise that it might have to do with the possible formation and persistence of electron density gradients owing to the presence of negatively charged dust (Kelley et al. 1998). For detection with L–band radar at $\sim$1 Ghz, the Bragg condition for radar backscatter requires density irregularities with scale size $\sim 1.5 \cdot 10^{-1}$ m.

The very high altitude trail observations with radar reported by Pellinen–Wannberg and Wannberg (1994; 1996) and Brosch et al. (2001; 2004) were performed during or close to solar cycle maximum conditions (1990 and 1999), during which the atmospheric and ionospheric densities usually expand to higher altitudes (Fig. 1). The possible connection between high altitude trails and solar cycle should be carefully studied during the approach of the next solar maximum within the coming years (Pellinen-Wannberg et al. 2010).

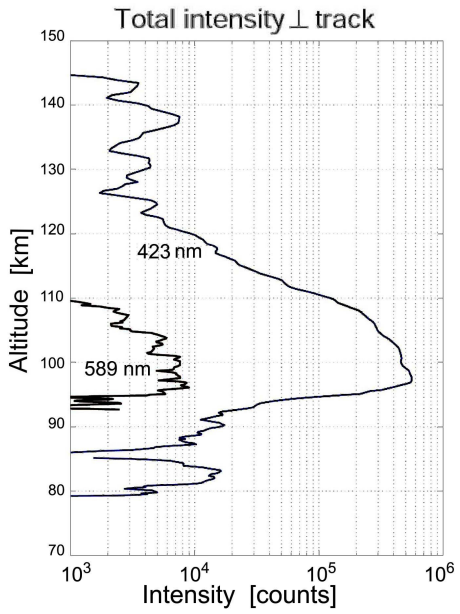


Fig. 4 Two optical altitude profiles of the same meteor through a 589 nm filter and a 423 nm filter. While the profile at 589 nm is quite typical for the usually quite strong sodium emission, the profile measured with the 423 nm filter starts at much higher altitude and is much stronger than expected for calcium or iron emissions within the filter bandwidth. The line is suggested to originate from the H_γ line at 433 nm associated with strong volatile H_2O gas emission (Pellinen-Wannberg et al. 2004).

3.5 Non-thermal ablation

Sputtered particles may cause formation of ionized meteor trails recorded by radars and may produce observable (high altitude) meteors, even if the mass loss due to sputtering is small. Sputtering is the only ablation mechanism for small particles, which are decelerated before being substantially heated. The sputtering of meteoroid material and the scattering of atmospheric species at meteoroids are connected to the formation of high altitude meteors (section 3.4). Sputtering generally occurs for impacting ions, but also for other particles (for instance neutral atoms, molecules) and it occurs over a large range of energies from sub - keV to MeV (we refer the reader to Behrisch and Eckstein 2007). A couple of percent of the sputtered species are ions and the removed atoms or ions are not necessarily in their ground states.

Popova et al. (2007) have suggested that the oxygen luminosity observed at high altitudes can be generated by the interaction of the sputtered atoms and ions with the background ionosphere (several percent of sputtered species are ionized). The sputtered atoms and ions typically have kinetic energy more than 1000 times the energy of surrounding ionospheric particles. When thermalizing through collisions they generate the luminosity. This is shown by Vinkovic (2007) in a model that also predicts the spatial extension of the generated luminosity. Rosenberg (2008) suggests that fast ions sputtered from a meteoroid might generate waves in the background plasma possibly detectable by radar scattering.

The sputtering models used in the studies mentioned above have limitations. The conditions for sputtering of meteoroids entering the atmosphere are determined by the atmospheric density (see section 2) and the entry speed. Assuming a maximum meteoroid entry speed of 72 km/s and impact of atmospheric O^+ one obtains a maximum energy of ions of 430 eV, corresponding to 27 eV per nucleon. The major species impinging on the entering meteoroids are O, O_2 and N_2. For meteoroids with entry velocity smaller than about 40 km/s energies of impinging atmospheric species are close to the threshold energies for sputtering and those are in most cases determined with great uncertainty.The sputtering yield at sub - keV energies can be influenced by many processes. Precise composition and structure of meteoroids as well the state of their surfaces are unknown. There appears to be some experimental evidence that sputtering from oxygen covered surfaces of various elements can produce high ion fraction (see Fig. 2 in Krauss and Grün 1979). The meteoroid surface may indeed be oxygen covered, since it is traveling through the ionosphere. Also atmospheric species may react at and be scattered from the meteoroid surface.

Opposite to what was recently suggested by Rogers et al. (2005) and Hill et al. (2005), sputtering does not cause significant mass loss. Kero (2008) reconsidered the mass loss using a numerical model of sputtering and thermal ablation similar to that used by the former authors. He suggested that the atmospheric densities in the Rogers et al. (2005) paper are a factor of 100 too high and that this led the authors to overestimate the mass loss caused by sputtering (note that the atmospheric densities are not explicitly given in the Rogers et al. (2005) paper). Kero found that for a 10^{-6} kg test particle with initial velocity 60 km/s sputtering of meteoroid mass is negligible compared to (conventional) ablation. Vondrak et al. (2008), based on another ablation model, find that sputtering only causes significant mass loss (35 percent at maximum) for the smallest meteoroids ($< 10^{-14} kg$) with high initial velocities ($< 45\ km\ s^{-1}$).

3.6 Fragmentation events

Numerous observed meteor phenomena are thought to result from fragmentation events. Without considering fragmentation, the meteoroid masses estimated based on meteoroid deceleration (dynamic mass) are largely different from those derived from the observed radiated energy (photometric mass). Radar reflections from meteor trails often differ from the predictions of simple models. Fragmentation also causes interference patterns in head echoes.

Currently two basic approaches are followed to describe fragmentation.The dust ball model assumes that meteoroids consist of individual dust particles that are released during fragmentation (Hawkes and Jones 1975). The Quasi–continuous fragmentation model (QCF) assumes a gradual release of the smallest fragments from the surface of a parent meteoroid and their subsequent evaporation, without specifying the fragmentation mechanisms (Bronshten 1983).

Hawkes et al. (2004) described high (spatial and temporal) resolution optical observations with the dust ball model and suggested that fragmentation occurred for at least half of the studied very faint meteors. Also light curves of faint Leonids ($\sim 10^{-7}$– 10^{-4} kg) and other meteor showers are quite reasonably explained in the frame of the dustball model (Koten et al. 2004). An ablation model for faint meteors that included fragmentation similar to the dust ball model was elaborated by Campbell–Brown and Koschny (2004). Campbell-Bown and Close (2007) further modified the same ablation model to calculate ionization densities; they studied 25 head echoes in detail and found that most of the observations

are consistent with meteoroid fragmentation, though their was no evidence for large spatial separation of the fragments.

Some of the observations of fragmentation events suggest that there is no large separation between fragments and a large fraction of meteor phenomena can be reasonably reproduced without taking fragmentation into account, Considering fragmentation is nevertheless important for analyzing the observations, since the derived material density of the meteoroid depends strongly on assuming or not assuming fragmentation (Bellot Rubio et al. 2002). Babadzhanov (2002) successfully applied the QCF model to 111 our of 197 observed bright photographic meteors (m $>10^{-5}$ kg) to derive mean bulk densities of meteoroids belonging to six meteoroid streams and to the sporadic background. He points out that the derived bulk densities are in average larger by one order of magnitude than proposed previously not taking into account the fragmentation.

3.7 Atmospheric dust measurements

Entering meteoroids leave layers of metallic atoms and some metallic ions in the ionosphere. As reported by Krankowsky et al. (1972), the reported isotope abundances of metals seem to conform with solar isotope abundances. LIDAR measurements have measured such layers consisting of Na, Fe, K, and Ca (McNeil et al., 2002). The metal layers are maintained in a steady-state by the equilibrium between constant dust entry followed by vaporization and removal via complex chemical reactions between the metal atoms and the atmospheric species in their descent to the ground (see Plane, 2002). Besides the neutral elements, ionic species, such as Fe^+, Mg^+, Ca^+ and other minor metallic constituents are also generated in the ablation of the dust particles. Metal ion formation from the atoms is mainly due to charge exchange reactions and to some extent due to collisional excitation, less so due to photo-ionization (see e.g. Plane, 2002).

Rosinski and Snow (1961) first introduced the idea that meteoric vapors recondense into nm-sized particles (i.e. smoke) (corresponding to a few 100s of silicate molecules) at altitudes near 80 –90 km. Hunten et al. (1980) later calculated the evolution and fate of these particles based on the assumption that the particles condensed to sizes of 1 to several 10s of nm; they considered the evolution of the particles taking into account coagulation, diffusion, and sedimentation. At 90 km altitude, the density of 20 nm particles thus formed exceeds that of the incoming interplanetary dust. Kalashnikova et al. (2000) calculated the formation of the meteoric smoke in a similar approach and found that the smoke formation depends on the properties of incoming micrometeoroids. The altitudes of the meteoric smoke particles (80-100 km) are also the altitudes, at which polar mesospheric summer echoes (PMSE) and noctilucent clouds occur. The former refers to the strong radar echoes at 50 MHz-1.3 GHz from the summer polar mesosphere at altitudes of 80 – 90 km (Czechowsky et al., 1979; Cho and Rötger 1997), while the latter are luminescent clouds observed shortly after sunset slightly above 80 km. Both phenomena are associated with the presence of meteoric metals, which catalyze the formation of ice crystals, since at these altitudes homogeneous nucleation is too slow as a source for ice crystals. Megner et al. (2008) have shown that transport of meteoric material might affect the coagulation conditions. The composition of the meteoric smoke is also important; for example sodium bicarbonate ($NaHCO_3$) is particularly effective as a condensation nucleus and has been suggested to facilitate the formation of noctilucent clouds (Plane 2000).

In 1992 Schulte and Arnold reported in-situ detection of meteoric smoke particles with mass spectrometer measurements. Kelly et al (1998) reported on the simultaneous radar

and rocket detection of a possible meteor condensation trail and meteoric dust in the upper mesosphere. The size of the dust was estimated to be $\sim$ 50 nm based on the fall speed. Since then the meteoric smoke particles have been studied by several dedicated rocket-borne experiments and experimental results are shown in Fig. 5. The detection techniques are further discussed in Section (6.1. The experimenters report the detection of both negatively - and positively - charged dust (e.g. Havnes et al. 2001; Rapp et al. 2003). In the tropical mesosphere a layer of positively charged nano–particles was found at altitudes $\sim$90–100 km (Gelinas et al. 1998). The concentration of these particles appears to be lower than in other experiments ($\sim 10^7$ m^{-3}). Particles at the bottom of the layer were charged negatively.

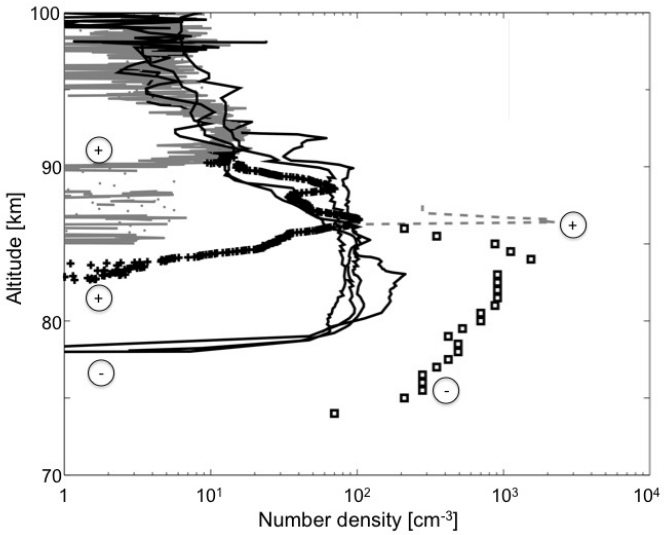


Fig. 5 The reported in-situ measurements of meteoric smoke in the ionosphere. The data are from Gelinas et al. (1998, grey solid line), Horanyi et al. (2000, grey dashed line and squares), Lynch et al. (2005, black solid line), and Rapp et al. (2005, crosses). Grey solid line, grey dashed line and crosses denote measured positive charges and the black solid line and the squares denote detected negative charges. The measurements are further discussed in the text and this figure is adapted from Megner (2008)

Mitchell et al. (1995) and Croskey et al. (2001) sampled both light (2200 amu) and heavy (20000 amu) positively charged particles with concentration 10^8–10^{10} m^{-3} during the rocket flights at altitudes 80 – –90 km ($3\cdot 10^{-24}$ – $3\cdot 10^{-23}$ kg, i.e. particles 1–6 nm in size assuming density 10^3 kg/m^3). These observations were performed during the noctilucent cloud season at high latitude, so it remains unclear whether these particles represent meteor smoke or water vapor condensation products. During later flights in Alaska a layer of negatively charged particles with masses 5000–10000 amu and concentration $\sim 10^8$ m^{-3} ($8\cdot 10^{-24}$– $2\cdot 10^{-23}$ kg) were observed (Lynch et al. 2005; Gelinas et al. 2005). The authors suggest that these charged particles represent 10–20% of the total number of dust particles within this size range (we will further discuss dust detection in Section 6). These latter flights were accompanied by lidar observations of Na and Fe atom layers and the observations suggested correlation between charged particle distribution and distribution of the iron atoms.

3.8 Possible instabilities in the Earth' s ionosphere

In the ionosphere, charged dust occurs naturally in the meteor ablation zone ($\sim 80-120$ km) and in particular in polar mesospheric regions such as NLC and PMSE regions ($\sim 80-95$ km altitude). At these altitudes, Debye lengths can range from about centimeters to millimeters, depending on plasma conditions such as daytime or nighttime, as well as factors such as precipitating particle fluxes. Estimates for the number density of small dust particles in the ionosphere, of size $\sim$ a few tens of nm to sub–visual, appears generally to be larger than about 10^8 m^{-3} under PMSE conditions (Havnes et al. 2001), so these plasmas may be considered as "dusty plasmas".

Although it has long been thought that dust can form from ablated meteoroid material (Rosinski and Snow, 1961; Hunten et al, 1980), little is known about the sizes or number density of dust in meteor trails. To get an idea of a limit on the dust density in a meteor trail, let us assume that when a meteoroid ablates, all the material condenses and coagulates into dust of size 5 nm. If the meteoroid is initially large, say 10^{-2} m, the average density of dust in a cylindrical trail of size 40 m by 10 km would be on the order of $\sim 10^{10} m^{-3}$. Since this value could be comparable to the plasma density in the trail, the dust, if charged, might affect the collective behavior of the meteor trail plasma including the behavior of waves and instabilities. Note that charged dust in a meteor trail may affect trail diffusion. It has been suggested that the presence of negatively charged dust may slow down the diffusion by reducing electron diffusivity, similarly to the possible effect of negatively charged dust in preserving plasma irregularities in PMSE regions (Kelley et al. 1998; Zhou and Kelley 1997). Kelley (2004) presented a new explanation for long duration meteor trains involving persistent charged dust trains. It has also been suggested that the effects of background dust on the decay time of underdense meteor trails might provide information on the subvisual smoke particles that are thought to exist (Havnes and Sigernes 2005).

In the upper D –lower E regions of the ionosphere which includes the altitudes of PMSE and the meteor ablation zone where trails form, the electrons have small gyrofrequency compared to their collision frequency (they are magnetized) while the ions are unmagnetized with collision rates larger than their gyrofrequency. In the lower E region, electron $\mathbf{E} \times \mathbf{B}$ or cross-field drifts can drive the known electrojet instabilities, such as the Farley-Buneman instability. These kinds of ion wave instabilities have been observed by radar returns from both equatorial and auroral electrojet regions, as the excited waves satisfy the Bragg scatter condition that the wavelength is 1/2 the radar wavelength (see e.g. Kelley 1989).

It has been suggested that these types of instabilities may apply to the formation of irregularities in meteor trails (Chapin and Kudeki 1994; Pellinen-Wannberg and Wannberg 1996). Oppenheim et al. (2003) and Dyrud et al. (2002) showed that Farley-Buneman type instabilities could occur in meteor trails, driven by electron cross-field drifts due to plasma gradients or ambipolar electric fields in the trail, and that the instabilities can lead to non-specular meteor trail echoes. Thus, it is of interest to understand how charged dust can affect these and related wave instabilities in the lower ionosphere.

Charged dust may affect the Farley-Buneman instability (Rosenberg and Chow 1998). As discussed above, dust can affect the phase speed of ion waves. At altitudes < 95 km, the phase speed increases as the negative charge density of the dust increases, and in analogy with ion acoustic waves, the critical electron drift decreases so the Farley-Buneman instability could be more easily excited.

The possibility of ion acoustic instability in dusty meteor trails in the low E region was considered by Rosenberg and Merlino (2007). When $T_e \sim T_i$ (often the case in ionospheric plasmas), the critical electron drift, parallel to $\mathbf{B}$, for an ion acoustic instability is roughly

$\sim$ the electron thermal speed. However, if the dust charge density is comparable to the ion density n_i, the ion acoustic speed increases, ion Landau damping decreases, and the critical electron drift decreases, as discussed in section (2.4). Rosenberg and Merlino (2007) considered an expanded trail (radius 10 meters, length 10 km) at an altitude $H \sim 95$ km, with $n_i \sim 10$ times larger than nighttime background plasma density, containing negatively charged dust of size $\sim 1/2$ nm with $Z_d \sim 1$ and $n_d \sim n_i$. It was found that the critical electron drift could be reduced by $\sim 1/4$ when $n_i/n_e \sim 10$, implying parallel electric fields on the order of 10-20 mV/m for ion acoustic wave excitation. Perhaps this mechanism may have some relevance to the observations of enhanced ion acoustic echoes in meteor trails reported by Pellinen-Wannberg and Wannberg (1996) but further work is needed to scope out the range of dust and meteor trail parameters required for instability.

The possibility of a very low frequency dust acoustic instability in a dusty meteor trail in the upper mesosphere was considered by Rosenberg and Shukla (2000; 2002). A relatively large dust charge is favorable for dust acoustic instability, but even larger dust of size tens of nm would get charged negatively to only a few electron charges in the cold ionosphere. Rosenberg and Shukla (2000) suggested that if the dust has low work function (< 4 eV), grains could get charged positively to larger charge states by photoemission in daytime conditions (see also Havnes 1990). Assuming positively charged dust with charge density comparable to n_i, it was shown that a dust acoustic instability might occur in an expanded trail (radius $\sim$ tens of meters), driven by an electron $\mathbf{E} \times \mathbf{B}$ drift of speed less than the ion thermal speed (Rosenberg and Shukla, 2000). The parameters corresponded to a large meteor, with estimated dust mass in the trail $\lesssim 100$ g. Also D'Angelo (2003) discussed a possible cross-field dust acoustic instability in the E region where the ions as well as the electrons are collision dominated.

D'Angelo (2005) considered whether ion acoustic and dust acoustic instabilities driven by electron or ions streaming along a vertical electric field in the high latitude $H \sim 85$ km ionosphere could be a cause of PMSE. D'Angelo (2005) points out that some observed features of PMSE (such as an anti-correlation of electron fluctuations relative to ion and dust fluctuations) point toward dust acoustic waves as a possible cause of the strong radar returns, but large electric fields (~ 1 V/m) would be required to excite the waves. Observations of large electric fields in these regions, up to 1 V/m, have been reported (Zadorozhny et al. 1993) although there is controversy regarding this value, with other experiments finding fields $\sim$ a few tens of mV/m (e.g. Holzworth et al. 2001). Scales and Ganguli (2004) also considered dust acoustic instability as a possible cause of PMSE, with the instability driven by ion or electron flows in the boundary of localized dusty plasma clouds in the mesosphere.

Dusty plasma instabilities at higher altitudes in the upper E – lower F region have been investigated in relation to radar scattering from rocket exhaust plumes. Bernhardt et al. (1995) suggested a dusty plasma mechanism, among others, to explain the enhanced radar backscatter from space shuttle exhaust observed with the Arecibo 430 MHz radar. The mechanism involves the condensation of water vapor exhaust into ice dust particles that get charged, leading to an instability (e.g., dust acoustic instability) due to the streaming of the charged dust relative to the background plasma. This could lead to radar scattering when the excited waves satisfy the Bragg scattering condition. Later, Rosenberg et al. (1999) considered a lower hybrid (LH) instability driven by a beam of negatively charged dust streaming across $\mathbf{B}$, finding that this instability could be relevant for explaining the enhanced radar backscatter from space shuttle exhaust. Rosenberg and Sorasio (2006) have considered the possibility of LH instability in dust-gas formations from rocket exhausts injecting combustion products into the ionosphere (see Platov 2004).

4 Dust in the interplanetary medium

Dust grains in interplanetary space are affected by impacts of the solar wind ions which cause sputtering. Some of the impacting ions become implanted in the dust and can be released as neutrals. Dust-dust collisions lead to fragmentation and vaporization. The vaporized material as well as the released neutrals become ionized and form a subpopulation of the solar wind pick-up ions. The dynamics of the dust is determined by gravity and the Poynting-Robertson force. The Lorentz force, which for the larger grains is a small perturbation, becomes dominant for the small grains, in particular the nano dust, which can be accelerated by the Lorentz force to velocities comparable to the solar wind velocity. Such fast nano dust was recently discovered by the STEREO spacecraft.

After describing dust destruction and dust interactions (section 4.1) we consider the evolution of dust generated species in the solar wind and shortly address the connection of dust to the measured pick-up ions, the neutral solar wind and the interplanetary field enhancements (section 4.2). We then discuss the dust dynamics and elaborate the acceleration of nano dust in the solar wind (4.3).

4.1 Dust destruction and dust interactions

A fraction of the impinging solar wind leaves the dust after interaction. The penetration depth Δ of solar wind protons into dust material is of the order of several 10 nm, the exact depth depending on the dust material. Ions that pass through the dust particles with radius $R < \Delta$ generally leave the dust in different charge state, often neutral (Minato et al. 2004). The momentum transfer onto the dust leads to the plasma Poynting-Robertson effect (Section 4.3). The solar protons and other solar wind ions impinging on large ($R >> \Delta$) dust particles remain in the surface layer. Diffusion of the implanted species, recombination and chemical reactions may occur. The outer layers of the dust grains that remained in the interplanetary medium for a couple of hundred years are supposedly saturated with solar wind particles, similar to the lunar regolith (Banks 1971). Fahr et al. (1981) suggested that solar wind particles impinging on such a saturated dust surface recombine and cause the trapped species to be released as neutrals again (”desorption”). Impinging particles also deposit heat, sputter ions out of the dust, damage the lattice structure and modify the mineral composition.

An important process that releases dust material is sputtering. Mukai and Schwehm (1981) considered sputtering of cosmic dust analog materials magnetite (crystalline structure of Fe_3O_4) and obsidian (glassy material consisting to 2/3 of SiO_2 and of MgO and Fe_3O_4) by protons and α particles and found that the dust material is predominantly sputtered by impinging protons of energies beyond typical solar wind energy, in majority by solar energetic particles. The sputtering rate is higher for He and hence increases with increasing He abundance. For the dust species and interplanetary medium conditions considered Mukai and Schwehm find that the dust particles suffer only moderate erosion by sputtering before reaching the dust sublimation zone (see Section 5.2.5 related laboratory studies). Dust near the Sun can also be destroyed by rotational bursting (Misconi 1993, Abbas 2004). The dust PR effect varies and is enhanced during coronal mass ejections (Ragot and Kahler 2003).

The dust and meteoroids are subject to collisional fragmentation, which may account for their size distribution in the interplanetary medium (Dohnanyi 1978). The dust collisional fragmentation and vaporization processes were discussed by Tielens et al. (1994) and Jones et al. (1996) . The impact of the dust particle generates a shock wave in the target grain. The

evolution of the shock wave within the particles can be studied in analogy with the corresponding wave in the bulk solid, with the experiments, although mostly restricted to lower collision velocities (<km/s) and larger projectile sizes (mm and cm) providing additional information. In this way the authors obtained some results concerning the distribution of fragments and the vaporized mass which were applied to dust collisions in the interstellar medium. Vapour production requires high collision velocities (>10 km/s) and in the interplanetary medium it usually occurs in the inner solar system (Mann and Czechowski 2005).

Neutral species are quickly ionized in the interplanetary medium (see Section 4.2), but a fraction of the forming vapor is already ionized. Calculations by Hornung with a hydrodynamics code predict that for typical impact velocities the vapour is in neutral charge state and that singly or doubly ionized species only form for impact velocities of the order of 100 km/s (Hornung et al. 2000). Charge production is, however, already observed for lower speeds, of the order of few km/s. Laboratory experiments at single charged grains accelerated in an electrostatic accelerator show that the charge released upon impact increases with the impact speed roughly as

$$Q \simeq 0.7 m^{1.02} v^{3.48} \tag{2}$$

where charge Q is in Coulombs, mass m is in kg and impact speed v is in km/s. There is a deviation from this relation for small velocities, where charge production is higher (McBride and McDonnell 1999).

Dust particles sublimate when they reach vaporization temperature. Describing dust sublimation in space requires extrapolation from the results obtained under laboratory conditions. Though the conditions vary with size and composition, sublimation becomes important within 0.1 AU (about 20 solar radii) from the Sun. The sublimation distances cannot be derived from observations. A canonical value of a dust free zone radius around the Sun is about 4 solar radii where most dust compounds sublimate, even though calculations show that some dust particles may survive further inward (see Mann et al. 2004 for a review).

According to a dust fragmentation model (Mann and Czechowski 2005) the majority of ions that are generated from the dust in the region outside roughly 0.1 AU from the Sun are produced during dust collisions. The production rates of pick up ions between 0.1 and 1 AU are (0.14 - 1.2) $\cdot 10^3$ kg m^{-3} from solar wind desorption, <0.14 $\cdot 10^3$ kg m^{-3} from solar wind passing dust and (0.17 - 36) $\cdot 10^3$ kg m^{-3} from collisional vaporization (Mann and Czechowski 2005). The pick up ions production from sungrazing comets is at average 0.16 $\cdot 10^3$ kg m^{-3} and its flux near Earth peaks between August and January (Bzowski and Krolikowska, 2005).

4.2 Dust generated species in the solar wind

Once the molecular and atomic species are released they are ionized; the major ionization mechanisms are photo ionization, charge exchange ionization with solar wind protons, and electron impact ionization close to the Sun (Mann and Czechowski 2005). The neutrals released inside of 1 AU are ionized on time scales small compared to the dust orbital periods. The ionized atoms or molecules are then subjected to Lorentz forces and gyrate about the magnetic field lines transported with the solar wind. This was studied in detail by Vasyliunas and Siscoe (1976). The ions are picked up by the solar wind within the Larmor rotation time of the order of seconds. (The Larmor rotation time for a proton at 1 AU is 300 s and inversely proportional to q/m). The ion initial velocity distribution ranges from zero to twice the solar

wind speed in the resting frame. It evolves as result of pitch angle scattering within the solar wind. Measuring the velocity distribution of pick-up ions allows to infer the distance from the Sun, where the ions were generated. The interstellar gas that enters the solar system provides the major component of neutrals in the interplanetary medium and also the major source of pick up ions.

If released as neutrals, then the dust-generated ions that are carried in the solar wind are predominantly singly charged, except for those that are released very close to the Sun. Species that are generated by dust destruction in the inner corona where the charge states of solar wind ions are determined by electron collisions may not be distinguished from the solar wind. Lemaire (1990) hence considered the infall of heavy ions from dust into the solar corona and its possible influence on the solar wind composition.

A part of the Solar Wind Ion Composition Spectrometer (SWICS) experiment aboard the Ulysses spacecraft comprises a detailed study of the pick-up ions in order to quantify some interstellar neutral gas components that enter the solar system (see Gloeckler et al. 2010 and references therein). Aside from pick-up ions that are generated by interstellar neutrals, SWICS discovered the pick-up ions associated with a so-called "inner source", which originates from the vicinity of the Sun and includes the elements H, He, C, N, O, Mg, Si, Ne, and molecular ions (see Gloeckler et al. 2010 and references therein). Attempts to explain the inner source with the dust surface scattering process suggested by Fahr et al. (1981) failed to explain the amount of detected ions (Mann et al. 2010a).

Dust related sources were also discussed in the context of neutral measurements, again without quantitative agreement (ibid.) A neutral atom released from dust, if ionized by charge exchange with the solar wind ion, gives rise at the same time to a pick up ion and a neutral solar wind atom. In consequence the flux of neutral solar wind hydrogen generated in this way should be similar to the total flux of inner source pick ions generated from dust. Collier et al. (2003) considered a similar mechanism, but assumed that the charge exchange occurs on the dust surface. He found that to account for the observed neutral solar wind flux a very large value of dust geometric cross section integrated between the Sun and 1 AU is necessary.

The mass loading by species released in cometary dust trails was also considered in the context of Interplanetary Field Enhancements (IFEs). IFEs are observed features in the time profile of the interplanetary magnetic field that seem neither instrumental nor correlated to other solar wind properties. Experimenters suggested their association to cometary dust trails or to other dust sources (Jones et al. 2003, Russell 2010). Mann et al.(2010a) estimated the charge production rate due to collisions within the dust trail and due to collisions of trail particles with interstellar dust and found that, with the current empirical models of vapour production, the collision rates are far too small to generate amounts of gas comparable to the solar wind densities.

4.3 Dust dynamics in the interplanetary medium

The main forces acting on dust in the interplanetary medium are solar gravity $\mathbf{F}_g$, radiation pressure $\mathbf{F}_{rad}$, solar wind pressure and the Lorentz force.

The radiation pressure force in the particle rest frame can be separated into the direct radiation pressure force and the velocity-dependent term that is referred to as the Poynting-Robertson (P-R) drag force. The latter causes the gradual loss of orbital angular momentum and orbital energy. The dust migrates toward the Sun. The momentum transfer from impacting solar wind particles, similarly to the radiation pressure force, can be decomposed into

direct solar wind pressure and the solar wind drag. Even though the force due to the plasma drag is much smaller in magnitude than the radiation pressure force, the corresponding drag on the orbital motion is comparable to the PR drag (plasma or pseudo Poynting-Robertson (PR) effect, e.g. Mukai and Yamamoto 1982; Minato et al. 2004). This results from the greater aberration angle than in the case of photon impact, since the solar wind speed is much smaller than the velocity of light. The PR lifetimes of dust within 1 AU are smaller than 10^6 years. The plasma PR effect varies with the solar cycle. It may have been more significant in the early solar system, which presumably had a stronger solar wind (Minato et al. 2006). Newly formed particles for which $\mathbf{F}_{\rm g} \sim \mathbf{F}_{\rm rad}$ can be ejected outward in hyperbolic orbits. The exact conditions depend on the dust initial velocity and on the value of the radiation pressure force. This is typically the case for dust in the size interval of a few 0.1 to several μm and these particles are called β-meteoroids (see Fig. 2).

Let's now consider the Lorentz force, the relevant parameter for which is the charge to mass ratio q/m, which is proportional to $1/R^2$ (R is the grain radius, deviations from this proportionality are discussed in Section 5). Assuming surface potential of $\sim$ 5 V, the charge to mass ratio for a dust particle with $\sim$1 μm radius is $\sim$5 10^{-10} e/m$_p$ (with elementary charge, e and proton mass, m$_p$). The corresponding Larmor frequency Ω_L at 1 AU can be found from the relation $\Omega_L \approx (q/m)10^4 \cdot$ B s^{-1} where q/m is in e/m$_p$ and the magnetic field B in Gauss. Taking $B \sim$30 μG gives Ω_L=1.5 $\cdot$ 10^{-10} s^{-1} which is much less than the orbital frequency at 1 AU. During orbital motion, such dust particles successively pass the regions of different field directions pushing them away from the initial orbital plane. Hence, the main effect of the Lorentz force on the $\sim$1 μm dust particles is the spread of the orbits in inclination caused by the sector structure of the interplanetary magnetic field (Morfill and Grün 1979). Lorentz force becomes more important for smaller particles. The dynamics studies of the dust with sizes of the order 1/10 μm are reviewed by Mann et al.(2004) and further references are given there.

The following discussion of the fluxes of smaller nano dust is based on a paper by Czechowski and Mann (2010). To study the trajectories of nano dust in the solar magnetic field we use Parker's model (Parker 1958)

$$B_r = \pm B_0 \left(\frac{r}{r_0}\right)^{-2} \qquad B_\phi = \pm B_0 \left(\frac{r}{r_0}\right)^{-1} \cos\theta \qquad B_\theta \equiv 0 \tag{3}$$

where r$_0$ = 1 AU, and $B_0 \simeq 2-4\times 10^{-9}$ T (20$-$40 μG). Except for the immediate vicinity of the Sun or the regions directly above the solar poles, this model provides a good approximate description of the average magnetic field. Near solar minimum, the fields in the northern and in the southern hemispheres correspond to opposite polarity. The dividing line is tilted relative to the ecliptic plane: this leads to the sector structure of the interplanetary field. The sectors with opposite polarity are separated by a thin current sheet which is warped due to the effect of the tilt. The thickness of the current sheet is of order of 10^5 km. Near solar maximum the field configuration is more complex. However, the simple picture of a tilted dividing line was found useful even in this case, although with a high $\sim$90 deg tilt.

We assume that the charge to mass ratio of a dust particle of $\sim$ 10 nm radius is approximately $q/m = 10^{-5} e/m_p$. At the distance r = 0.1 AU from the Sun the solar magnetic field $B \simeq 3\times 10^{-7}$ T gives for such a dust particle the Larmor frequency $\Omega_L = 3\times 10^{-4} s^{-1}$. The Larmor radius R$_L$ is equal to $|v'_\perp|$ / Ω_L where $v'_\perp$ is the velocity perpendicular to the magnetic field in the frame where the electric field **E** = 0. In relation to the inertial frame centred at the Sun, this frame moves at the velocity equal to the transverse component of the solar wind velocity with respect to **B**, corresponding to approximately 45 km/s at $r = 0.1$

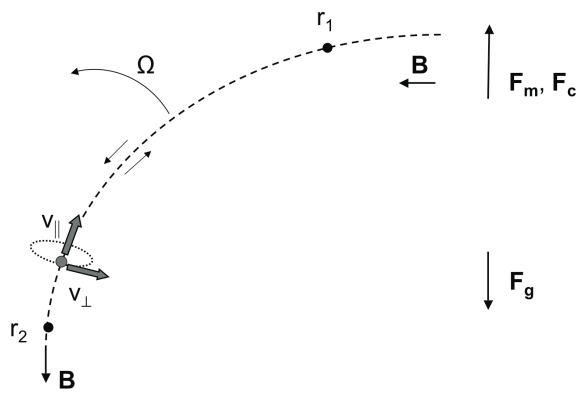


Fig. 6 Sketch of the nano dust trajectory in guiding center approximation. The freshly created dust particle is governed by solar gravity $\mathbf{F_g}$ directed inward, and centrifugal force $\mathbf{F_c}$ and mirror force $\mathbf{F_m}$, directed outward. It gyrates about the magnetic field line (dashed line) with the velocity $\mathbf{V}_\perp$ and follows magnetic field line rotation. A trapped particle oscillates along it. Particles are ejected when the combination of centrifugal and magnetic mirror force exceeds gravity.

AU. Assume that the dust particle is a fragment split off the larger dust particle moving at Keplerian velocity. The Larmor radius can then be seen to be $R_L = 1.5 \cdot 10^8$ m. The condition $R_L << r$ (and $R_L << L$ where L is the characteristic scale of the field) is therefore satisfied with a large margin. By similar reasoning, for $r = 0.3$ AU one obtains $R_L = 3 \cdot 10^9$ m, so that $R_L << r$ still holds. It follows that in a sizeable part of the inner ($r < 1$ AU) heliosphere the adiabatic condition $R_L << L$ is satisfied for $\sim$ 10 nm (or smaller) dust particles. The guiding centre approximation (e.g. Northrop 1963) provides then a good description of the dynamics.

The guiding centre velocity in the direction perpendicular to the field line is approximately equal to the component of the solar wind velocity that is transverse to **B**. In effect, the guiding centre motion consists approximately of sliding along the rotating magnetic field line. In Czechowski and Mann (2010) we found that the motion of the nano dust particle following from exact integration of the dust equation of motion can be understood qualitatively within the simple picture following from the guiding centre approximation. The rotation of the field line induces inertial forces, in particular the centrifugal force, acting in the direction opposite to that of the gravity force. The inertial force and the "magnetic mirror" force are independent on the value of q/m, implying that, in the region in which the guiding centre approximation works ($R_L << L$), the dynamics becomes approximately independent on the dust particle size.

Near the ecliptic, the centrifugal force overcomes gravity at about $r = 0.15$ AU. It can be seen that only the dust particles released at distances larger than this critical value do escape from the vicinity of the Sun and can reach 1 AU. The motion of the dust particles released at smaller distances can be understood as a bounded motion along the rotating field line between the initial (release) distance and the perihelium distance, where the particle is reflected by magnetic mirror.

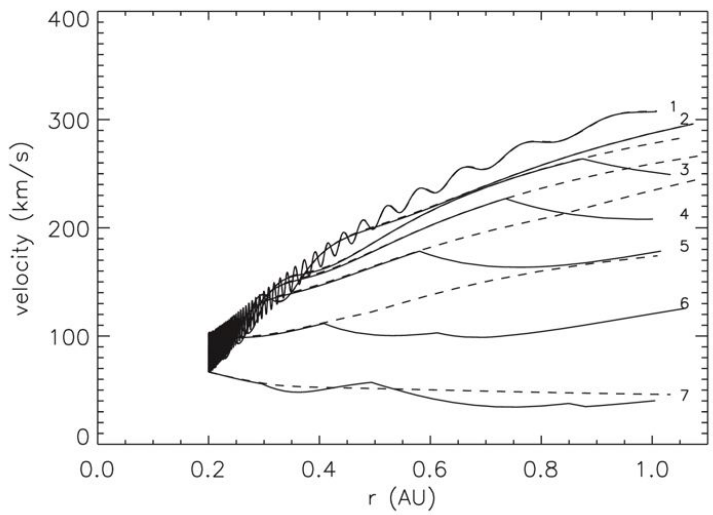


Fig. 7 The speed vs. distance for nano dust released at 0.2 AU (solid and dashed lines) from top to bottom with q/m =10^{-4}, 10^{-5}, $7 \cdot 10^{-5}$,$5 \cdot 10^{-5}$, $3 \cdot 10^{-5}$, and 10^{-5} e/m_p(from Czechowski and Mann 2010).

At 1 AU, a significant part of the $\sim$ 10 nm or smaller dust particles would come from the region near the Sun, where the collisional fragmentation rate of larger dust particles is high. Although at the distances approaching 1 AU the guiding centre picture ceases to be applicable to 10 nm size grains, the particles are are still approximately tied to rotating magnetic field lines. In result, their velocity at 1 AU is of the order of the velocity of the field line, which is of the order of the solar wind velocity. At the same time, these dust grains have an outward velocity component, acquired at smaller distances due to the centrifugal force (see Fig. 6). Fig. 7 shows the absolute velocity as function of distance for nano dust with different q/m released at 0.2 AU. The solid lines denote trajectories during focusing field configuration with incoming magnetic field in the northern hemisphere, so that the electric field -(1/c) V$\times$B points toward the current sheet. The dashed lines correspond to reversed antifocusing field configuration with outgoing magnetic field in the northern hemisphere and electric field pointing away from the current sheet. Trajectories with $q/m = 10^{-4} e/m_p$ and $q/m = 10^{-5} e/m_p$ are almost identical for the two field configurations and reach highest velocity. Acceleration quickly drops for smaller q/m (larger mass). Crossing the current sheet (sharp bend in the velocity profile) also decreases the effect of acceleration. Grains with q/m=10^{-7} (about 100 nm) are ejected with moderate speed. Dust with sizes $\sim$10 nm or smaller near 1 AU should therefore in majority stream outward with velocity $\sim v_{SW}$). The trajectory calculations combined with a dust collision model lead to an estimated flux of dust with mass m $> 10^{-20}$ kg of ϕ = 0.6 m$^{-2}s^{-1}$ (Czechowski and Mann 2010).

Fast nano dust particles have recently been detected by the wave instrument of the STEREO spacecraft. The measured flux is highly time variable (Meyer-Vernet et al. 2009b). The estimated average flux of nanoparticles detected by STEREO during the year 2007 is shown in Fig. 2. It is of the same order as the estimate based on trajectory calculations and the collision models (Czechowski and Mann 2010).

Both the nano dust and the larger β-meteoroids cross 1 AU in outbound orbits, but have distinctly different velocities. Fig. 8 shows the average absolute velocities as func-

tion of grain mass. Though the Ulysses dust instrument made most of its measurements beyond 1 AU we compare the calculated velocities with the Ulysses results. The detected β-meteoroids have velocities $\sim 40-80$ km/s (Wehry and Mann, 1999), comparable to the shown values for the dirty silicate asteroidal dust model. The Ulysses measurements cover masses$> 10^{-19}$ kg which is larger than the nano dust.

The trajectory calculations do not provide final explanation of the flux variations. The nano dust presumably forms by fragmentation of dust of several micrometer sizes and larger, and we are unable to explain the high variability of the flux with variations of the dust source. The variations in the extension of the trapping zone would lead to time variation of the flux near 1 AU, but possibly also other mechanisms come into play. Since the nano dust is accelerated continuously along its trajectory, it is likely that the time variations are connected to solar wind conditions in the inner solar system, rather than to conditions near the spacecraft. We note that further analysis of the STEREO/WAVES measurements shows, aside from the detected nano dust, a second class of dust impacts that are possibly due to β-meteoroids and interstellar dust (Zaslavsky and Meyer-Vernet, work in progress).

Because of its high rigidity we can not exclude that the nano dust enters the Earths magnetosphere. If this is the case, the mass input into the ionosphere below 400 km altitude is $\sim 2\cdot 10^{-6}$ kg s^{-1} or 55 kg / year, which is $\sim 10^{-6}$ or less of the meteoroid mass input. Nano dust impacts are also suspected to generate impact craters detected on a foil that was exposed to space within the Earth magnetosphere on the International Space Station ISS (Carpenter et al. 2005) (also shown in Fig. 2). The possible connection of the flux derived from the impact craters to the flux derived from STEREO results was not studied.

In spite of the large speed of the nano dust, the total momentum carried by the nano dust component is small. Based on the Czechowski and Mann (2010) flux estimate the average momentum of nano dust at 1 AU is $6\cdot 10^{-21}$ kg $m^{-2}s^{-1}$ and this is $\lesssim 10^{-6}$ of the solar wind momentum assuming solar wind mass density from table 2.4_2 and solar wind speed 400 km / s. The trajectory calculations described above assume fixed values of q/m. Comparison of the q/m values used in our calculation with the values obtained in dust charging calculations by Mukai (1981) and Kimura and Mann (1998) can be found in the figure 11 of Czechowski and Mann (2010). The inverse square size$1/R^2$ scaling of the charge to mass ratio changes in the small size limit, because secondary electron emission becomes more important near the Sun and moreover small particle effects arise for the charging processes. This is further discussed in Section 5.

5 Dust charging and laboratory experiments related to dust in space

In the interplanetary medium and in the ionosphere the dust charging is based on the same physical processes, though the relevant parameters cover quite different ranges. Laboratory studies concerned with dust charging and other interaction of dust with electrons, ions and photons require experiments with single dust grains. This is accomplished in several laboratories by using quadrupol traps. For most of the experiments high vacuum conditions are required. Experimental results mainly confirm theory, but show the importance of effects due to the dust size, dust shape, as well as surface reactions and in general deviation from theory for small sizes. The measurements of secondary electron emission (SEE) reveal a component of low energy backscattered electrons, which is not predicted by theory. The sputtering experiments suggest that sputtering from heavy solar wind ions impacting on dust is more important for sputtering rates than previously assumed.

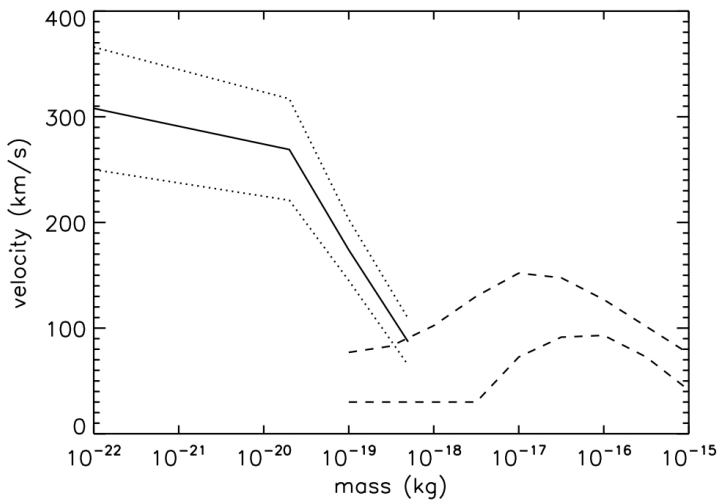


Fig. 8 The average velocity as function of mass at 1 AU for dust ejected by Lorentz forces (solid line and dotted lines) and dust ejected by radiation pressure (dashed lines). The shown solid and dotted lines are derived based on a model of nano dust production by collisional fragmentation inside 1 AU combined with trajectory calculations. The dashed lines show velocities are derived assuming that the dust is released from circular orbit at 0.1 AU and ejected outward by radiation pressure obtained for a highly absorbing cometary dust model (upper curve) and a dirty silicate asteroidal dust model (lower curve). This figure is adapted from Mann et al. (2010b).

This section is organized in a part describing the theory of dust charging (5.1) and in a part describing laboratory experiments to measure specific processes and the related numerical models (5.2).

5.1 Theoretical description of dust charging

A dust particle immersed in a plasma and a radiative environment collects charged plasma particles and may emit electrons and ions. In general, the incoming and outgoing electric currents do not balance each other initially, making the dust particle's electrostatic charge (and potential) vary, which in turn changes the currents. This proceeds until the dust particle's charge and potential set up so as to make the currents balance each other (see for example Whipple 1981, Mukai 1981, Goertz 1989, Mendis and Rosenberg 1994).

In the following discussions on dust charging, we consider first dust particles of size larger than the nm scale and approximate them as macroscopic spheres of radius R_d. We also assume that their separation is greater than the Debye length, that the velocity distributions of the plasma particles are Maxwellian, and only consider the collection of plasma particles and photoelectron emission. We describe the basics of dust charging in the solar wind and in the ionosphere (5.1.1) and then give the discuss the charging currents and electrostatic potentials in these cases (5.1.2) in order to finally describe the electric charge and charging time scale (5.1.3). More complex cases are considered at the end of this section (5.1.4).

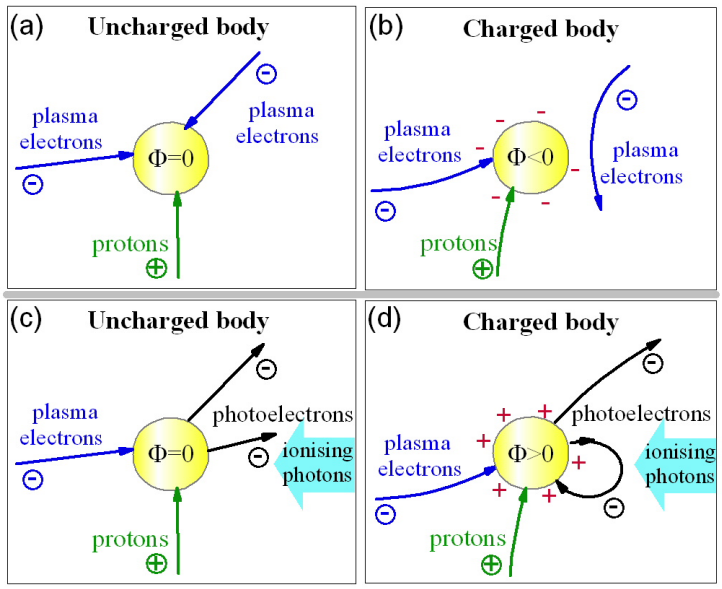


Fig. 9 Sketch of the charging of a body in a plasma (adapted from Meyer-Vernet 2007 page 353). In the ionosphere (top), the charging is governed by collection of plasma electrons and ions, with the former dominating because of their larger mobility (a), so that the dust particle is charged negatively until the electrostatic potential repels enough electrons to make the net current vanish (b). In the interplanetary medium (bottom), the flux of photoelectrons from an uncharged dust particle far outweighs the flux of incoming plasma electrons - itself much larger than the flux of plasma ions (c). This charges the body positively, until the electrostatic potential traps enough photoelectrons to make the net current vanish (d).

5.1.1 Some basics of dust charging

We begin the discussion by making order of magnitude estimates (Fig. 9).

Consider first the ionosphere. Photoelectron emission from the dust particle is generally small compared to collection of plasma particles, so that the charge is determined by the ambient electron and ion fluxes. Because of the very small electron mass, the electron flux is initially much greater than the ion flux, so that the dust particle charges negatively, until the negative charge repels sufficiently the incoming electrons to make their flux balance the ion flux. For doing so, the dust particle's electrostatic potential must ensure that the (positive) potential energy $-e\Phi$ outweighs the typical kinetic energy $k_B T_e \sim 0.03 - 0.2$ eV (depending of the altitude) of the plasma electrons. We deduce that dust particles charge to a negative potential of up to several tenths of volt in the ionosphere.

On the other hand in the interplanetary medium, an uncharged dust particle subjected to the solar ionising radiation generally ejects many more electrons than it collects from the ambient plasma, whereas the incoming ion flux is still smaller. Hence it charges positively, until this positive charge binds sufficiently the photoelectrons to make their flux balance that of the solar wind electrons. For doing so, the dust particle's potential Φ must provide the photoelectrons with a (negative) potential energy $-e\Phi$ that outweighs their typical kinetic

energy of a few eV. We deduce that a sunlit dust particle charges to a positive potential of several volts in the solar wind.

5.1.2 Charging currents and electrostatic potential

For dust particles much smaller than the Debye length and the collisional free path, the plasma currents are orbit-motion-limited (OML, see Laframboise and Parker 1973). Hence the electron and ion currents on a dust particle of surface $4\pi R_d^2$ at (negative) potential Φ in the ionosphere are

$$I_e = \pi R_d^2 n_e e v_e \exp\left(e\Phi/k_B T_e\right) \qquad \Phi < 0 \tag{4}$$

$$I_i = \pi R_d^2 n_i Z e v_i \left(1 - Ze\Phi/k_B T_i\right) \quad \Phi < 0 \tag{5}$$

since they are respectively repelled and attracted. Here Ze is the ion charge, $n_{e,i}$ and $v_{e,i} = (8k_B T_{e,i}/\pi m_{e,i})^{1/2}$ are respectively the electron, ion number density and average speed, so that for a given infinitesimal surface element, $n_e/2$ electrons per unit volume are approaching from one side with an average perpendicular velocity $v_e/2$. The (negative) equilibrium potential Φ is obtained by balancing the currents as

$$Ze\Phi/k_B T = -\ln\left[(m_i/m_e)^{1/2}(T_e/T_i)^{1/2}/\left(1 - e\Phi/k_B T_i\right)\right] \quad \text{ionosphere} \tag{6}$$

For example, with one singly charged ion species of mass $m_i \simeq 30 m_p$ and temperature $T_i \simeq T_e \simeq T$, the dust particle potential is $\Phi \simeq -3.9 k_B T/e$ volts.

In contrast, in the interplanetary medium, the dominating current is generally the photoemission current due to solar ionizing radiation. For most materials, the yield, i.e. the number of ejected photoelectrons per absorbed photon, becomes significant for photon energies above several eV, with a maximum yield in the range 0.05-0.5, depending on the physical and chemical structure of the dust particle. The photoemission current from an uncharged dust particle is found by integrating over the solar spectrum the product of the yield by the absorbed flux. It is in the range (Grard 1973; Mukai 1981)

$$I_{ph0} \simeq \delta \cdot 10^{14} e R_d^2 / r_{\mathrm{AU}}^2 \qquad \delta \sim 0.6 - 6 \tag{7}$$

where r_{AU} is the heliospheric distance in astronomic units and the smaller value of δ corresponds to low-yield materials such as graphite, the larger to high-yield photoemitters such as metals and silicates. Photoelectrons are ejected with a typical energy of 1-3 eV, which may be approximated by a temperature

$$T_{ph} \sim 1 - 2 \times 10^4 \ \mathrm{K} \tag{8}$$

This holds for an uncharged surface. For a charged surface whose potential Φ with respect to the distant undisturbed plasma is positive, the escaping photoelectrons are those emitted with a kinetic energy greater than $e\Phi$. Integrating over the velocities, this yields

$$I_{ph} = I_{ph0}\left(1 + e\Phi/k_B T_{ph}\right)\exp\left(-e\Phi/k_B T_{ph}\right) \qquad \Phi > 0 \tag{9}$$

The dust particle potential settles so that this current is balanced by the current of incoming plasma electrons (attracted)

$$I_e = \pi R_d^2 n_e e v_e \left(1 + e\Phi/k_B T_e\right) \quad \Phi > 0 \tag{10}$$

We have neglected for simplicity the solar wind proton contribution, which is smaller than the electron one. The dust particle equilibrium potential is thus given by

$$\Phi \simeq \left(k_B T_{ph}/e\right) \ln\left[\frac{5\cdot 10^9\ \delta}{n_{\mathrm{m}^{-3}}\ T_e^{1/2} r_{\mathrm{AU}}^2}\frac{1+e\Phi/k_B T_{ph}}{1+e\Phi/k_B T_e}\right] \tag{11}$$
$$\sim 1-10 \text{ volts} \quad \text{interplanetary medium}$$

where we have substituted typical solar wind parameters, and neglected the weak radial variation of $T_e^{1/2}$.

5.1.3 Electric charge and charging time scale

The electrostatic charge q on a dust particle of radius $R_d \ll L_D$ is thus

$$q \simeq 4\pi\varepsilon_0 R_d \Phi \tag{12}$$

With the typical potentials found above, this corresponds to a number of elementary charges

$$q/e \sim \text{a few } 10^8 \times R_d \qquad \text{(ionosphere)} \tag{13}$$
$$q/e \sim \text{a few } 10^9 \times R_d \quad \text{(interplanetary medium).} \tag{14}$$

This amounts to values up to $\sim$ tens and hundreds electrons for a dust particle of size 0.1 μm in respectively the ionosphere and the interplanetary medium.

The above estimates are equilibrium values. The equilibrium time may be estimated by viewing the dust particle in the plasma as an electric circuit of capacitance $C \simeq 4\pi\varepsilon_0 R_d$ and resistance $R \sim 1/\mid dI/d\Phi \mid$, where I is the current flowing between the dust particle and the plasma in absence of equilibrium, so that $\tau \sim RC$. For each charging process contributing to a current I, the derivative is of the order of magnitude of $eI/k_B T$, where I and T are respectively the rate and temperature of the charging process that yields the greatest derivative. For example for a dust particle in the solar wind, the charging time scale is of order of magnitude

$$\tau \sim C k_B T_{ph}/\left(eI_{ph0}\right) \tag{15}$$
$$\sim \frac{5}{R_{d(\mu\mathrm{m})} r_{\mathrm{AU}}^2}\ \mathrm{s} \quad \text{(interplanetary medium)} \tag{16}$$

for a dust particle of radius $R_{d(\mu\mathrm{m})}$ at distance r_{AU} from the Sun. This yields a few tens seconds for a 0.1 m grain at 1 AU. Charging time scales are possibly different for the nanometric dust . Laboratory measurements indicate the importance of electron attachment in the charging of nano dust (see below) and this may be relevant to the charging of the meteoric smoke particles that are thought to exist in the mesopause (see Megner et al. 2006).

Let's consider charging of sub–nm dust that is thought to be generated in meteor trails. For example, to estimate if there is enough time for sub–nm dust particle to charge in a meteor trail, we use the time scale for dust charging by attachment of a single electron:

$$\tau_e \sim \left(\pi R_d^2 n_e v_e\right)^{-1} \tag{17}$$

(we neglect other charging mechanisms in this estimate). The electron density in the trail at time t can be obtained by dividing the line density of the meteor trail, α, by πr^2, where r is the radius of the meteor trail given by

$$r^2 = 4Dt + r_0^2. \tag{18}$$

Here r_0 is the initial radius that is described in section 3.2 and D is the ambipolar diffusion coefficient; D is about 1 m^2 s^{-1} at 85 km and 140 m^2 s^{-1} at 115 km. This equation applies to the usual meteor ablation zone (see e.g. McKinley 1961). Thus the charging time (for larger t) goes as

$$\tau_e \sim 4Dt/(\alpha R_d^2 v_e), \tag{19}$$

so τ_e is smaller for larger meteors at lower altitude. For example, consider a meteor of mass $\sim 0.1g$ and line density $\sim 10^{16}$/m that ablates and forms a meteor trail at an altitude of about 95 km, where $D \sim 10$ m^2/s. For a dust particle of radius R, and for $t = 5$ s, we find that $\tau_e \sim 0.5$ s, so it appears the dust particle could charge quickly.

5.1.4 Additional processes

The above estimates neglect some processes which may be important in particular cases.

Firstly, electrons that are energetic enough can penetrate into the dust particle and excite electrons within it, producing secondary emission of electrons. This effect may be significant in the interplanetary medium for some materials. In that case it makes the dust particle's electrostatic potential more positive (see for example Chow et al. 1995). Furthermore, the suprathermal electrons present in the interplanetary medium can make the dust particle potential multivalued, i.e. there are two stable equilibrium potentials of opposite signs for a given dust particle in a given environment, so that an infinitesimal variation in properties can yield jumps in potential and similar dust particles can have charges of opposite signs (Meyer–Vernet 1982).

Secondly, dust particles are generally not spherical, so that the equilibrium charge is greater than the value (12). In practice, for non-spherical dust particles of volume V, (12) still holds with the effective radius $R_d = (3V/4\pi)^{1/3}$, provided that the length-to-diameter ratio is smaller than about ten, but the charge may be much greater for long rods and fluffy dust particles (Auer 2007).

Thirdly, even plasma current collection can be more complex than the simple expression given by eq. (5) in environments such as the lower ionosphere because of the presence of different ionic species including negative ions (e.g. Rapp et al. 2005).

For very small dust particles, further processes come into play. In the interplanetary medium, the potential is limited by field emission if the surface electric field $\sim \Phi/R_d$ (where R_d is the dust particle's radius or the smallest value of the curvature radius for a non-spherical dust particle) is greater than the value enabling the dust particle's electrons or ions to tunnel towards the surface. For positive potentials, only ion emission can take place, so that the critical field strength $\sim 3 \cdot 10^{10}$ volts/m is reached only for dust particles smaller than a fraction of nm. However in that case, the electrostatic stress $\sim \varepsilon_0(\Phi/R_d)^2$ generally exceeds the maximum tensile strength against fracture (itself generally smaller than $\sim 10^9$ N/m^2 - the value for tektites or polycrystalline bulk solids) (see for example Hill and Mendis 1981), so these dust particles are generally disrupted before ion emission can take place. The sub–nm dust is also possibly ionized due to electron collision or can be charged by electron attachment (Watanabe 2006). Finally, one sees from (13) and (14) that dust particles smaller than a few tens nm and a few nm in respectively the ionosphere and the interplanetary medium may have a charge smaller than unity, so that charge fluctuations are comparable to the charge itself. In that case, the charging process becomes statistical (Watanabe 2006; see also Rapp et al. 2005; Draine and Sutin 1987). Note that photo emission and photo-detachment were recently discussed for meteoric smoke composed of metal oxides (e.g., Rapp 2009).

5.2 Laboratory experiments and related numerical models

While several laboratory experiments have been carried out to study the parameters of grain ensembles within a plasma, experiments at single grains are still a challenge. We start by describing the experimental set ups to manipulate the single particles (5.2.1) and after discussing the scattered measurements of different charging processes (5.2.2) report about the detailed studies of the secondary electron emission (SEE) process (5.2.3), of the dust exposed to ion beams (5.2.4) and of dust sputtering (5.2.5).

5.2.1 Experimental set ups

Most measurements at single cosmic dust analogs are carried out based on the quadrupole trap technique, which was developed in the 1950s for ion spectroscopy (Paul and Steinwedel 1956). The Dusty Plasma Laboratory at Marshall Space Flight Center, Huntsville, Al, USA uses a quadrupole trap at pressures $\approx 10^{-3}$ - 10^{-2} Pa. In this device the dust surface charge is derived from the electrostatic force in the trap indicated by the compensating dc potential applied to the balance electrodes in order to levitate the particle (Spann et al. 2001, Abbas et al. 2001, 2004). The group studied charging of Apollo 17 dust grains exposed by low-energy electron beams (Abbas et al. 2008). This device is also used for radiation pressure measurements and for studying grain alignment and grain rotation(see Abbas et al. 2003), which are beyond the scope of our discussion.

Since the laboratory measurements bear the difficulty of separating the effects of the remaining atmospheric species in the device from the measured effect, the Space Physics Laboratory at Charles University, Prague, has developed a 3D quadrupole trap that operates under ultra-high vacuum conditions ($\approx 10^{-7}$ Pa or better). In this device the grain charge is derived from the recorded oscillation frequency of the grain that is (due to the specific shape of the electrodes) proportional to its charge-to-mass ratio (Cermak et al., 1994). Special techniques are also applied to measure the grain mass, diameter, and capacitance (Cermak et al. 1995, Zilavy et al. 1998, Pavlu et al. 2004). Also Grimm et al. (2006) studied the charging of dust particles in a high vacuum quadrupole trap, in this case the grains were exposed to soft x rays. The quadrupole trap measurements are typically carried out at grains of several 100 nm size. The challenges in studying smaller grains lie in handling the small particles, measuring their size and reducing effects due to the remaining pressure in the device.

Experimenters in the Dusty Plasma Group at University of Colorado at Boulder followed a different concept in which particles are dropped through the plasma volume and the dust surface charge is derived from the voltage pulses measured at a Faraday cup at the bottom of the volume (see Walch et al. 1994, 1995). This method detects an ensemble of grains and results can only be interpreted only statistically.

5.2.2 Measurements of different charging processes

This section briefly reviews investigations of interactions that are not discussed in detail in the following. Several laboratory experiments have been carried out to study specific issues of the grain charging within a plasma. Barkan et al. (1994) studied the difference in charging of micron-sized grains in a dense dust cloud to the charge on an isolated grain. Walch et al. (1994, 1995) measured the charge on small grains of non-conducting materials in a plasma containing both thermal and suprathermal electrons. Sickafoose et al. (2001) studied the photoemission and triboelectric charging of single dust grains levitating upon a

metal surface. Triboelectric charging which is the transfer of charge via mechanical contact, is an important charging process for planetary regolith analogs, as are secondary electron emission and photo electron emission. Triboelectric charging is also suggested as one of the mechanisms that generate secondary charges in the rocket- borne dust detectors (see section 6.1). Abbas et al. (2006) presented laboratory measurements of the photoelectric yields of grains of silica, olivine, and graphite of 0.1–5 μm radii illuminated with a UV radiation at 120–160 nm wavelengths. Large grains (tens of micrometers) had an order-of-magnitude higher yields than that reported for planar samples. Submicron grains with sizes several wavelengths of the incident photons had yields close to or below the bulk value and so far there is no agreement with theoretical results. Small size effects are also important for the secondary emission of electrons (SEE), which gains importance for small grains.

5.2.3 Secondary electron emission experiments (SEE)

SEE effects are usually underestimated but SEE is important whenever the energetic (above 50 eV) electrons are present. Theoretical descriptions of SEE still follow Sternglass (1957) who established a formula hat describes the yield of the secondary electron emission, i.e. the number of ejected electrons per infalling electron, as a function of the energy of the primary electron beam. The description bases on measurements at large planar samples. This yield rises with the beam energy to a maximum (at 0.3–2 keV) and then drops down to zero. The height of the maximum depends on the material. It is about unity for metals and reaches several tens for some insulators. Electrons leaving the grain consist of two different populations: true secondary electrons and backscattered primary electrons. The energy distribution of true secondary electrons is limited to about 50 eV, whereas scattered electrons can reach energies comparable to that of the primary electron beam. Suszcynsky and Borovsky (1992) modified the Sternglass theory based on measurements at electron beam energy 2–30 keV and at different materials, in particular insulators and ices.

The secondary emission of electrons is very sensitive to the surface curvature (Draine and Salpeter 1979). The yield increases with incident angle because the main interaction region is closer to the surface. Thus particles with highly curved surfaces and small particles with size comparable to the beam have enhanced yields (Richterova et al. 2007). For very small grain size, the beam electrons are not captured inside the grain and the surface potential can reach rather large values (e.g., Pavlu et al. 2008, Beranek et al. 2010a).

The increase of grain charges for small sizes is a common feature that is observed for many different materials (Svestka et al. 1993, Richterova et al. 2006a, Richterova et al. 2007, Beranek et al. 2009). This was expected from theoretical studies who show that SEE yield increases with decreasing size and becomes very large for grains with size of the order of the primary electron penetration depth. The measured surface potentials, however, exceeds that results from SEE. Richterova et al. (2004) simulated numerically secondary electron emission from spherical dust grains and found that backscattered electrons have considerable influence on the equilibrium surface potential, especially for high primary energies. The authors have improved the model including more complex principles of the electron-solid interaction and using an advanced Monte Carlo technique (Richterova et al. 2006a). The results can be briefly summarized as follows: (1) the scattering of primary electrons inside the grain is critical to understanding the secondary electron emission, (2) the grain charge is determined not only by the SEE yield, but equally important is the energy distribution of secondary electrons, (3) the increase of the surface potential with decreasing grain size is predominantly caused by the increasing number of backscattered primary electrons, not by true secondary electrons. The model was successfully verified by laboratory experiments on

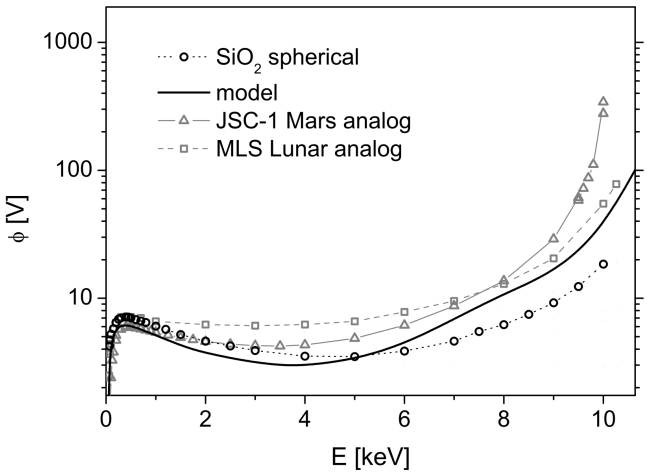


Fig. 10 Grain surface potentials, ϕ as a function of the primary electron beam energy for the JSC Mars-1 and MLS Lunar analogs together with the results calculated according the Richterova et al. (2010b) model for glass spheres with diameter of 0.6 μm and corresponding mass $0.25 \cdot 10^{-15}$ kg. The masses of measured grains are: SiO_2 – $1.3 \cdot 10^{-15}$ kg, JSC-1 Mars – $1.5 \cdot 10^{-15}$ kg, and MLS Lunar – $1.8 \cdot 10^{-15}$ kg.

spherical grains from different materials (Richterova et al. 2006a, 2006b, 2007, Beranek et al. 2009, Pavlu et al., 2009).

The measured surface potentials of Martian and Lunar analog grains are shown together with the results of the Richterova et al. (2006a) model in Fig. 10. The potential profiles roughly follow the profile of the secondary emission yield at low (below 1.5 keV) primary energies for all investigated grains. The masses and compositions of the grains are similar and the differences observed in higher energies can be attributed to the shape effect. To show it, we used the experimental measurements for SiO_2 spheres (Richterova et al. 2007) because this component is a principal constituent of both simulants. A comparison shows that the grain from the Lunar simulant behaves approximately as the SiO_2 sphere of a smaller (approx. by a factor of 2) size. It suggests that a better approximation of the Lunar sample shape would be an ellipsoid of revolution with the ratio of axes of about 1/2 and even smaller ratio would be appropriate for the Martian sample. A similar shape effect was found in Pavlu et al. (2008) where the charging of grain clusters was widely discussed. The positive grain potential can reach several hundred of volts and it is limited by the threshold of the ion field emission.

Motivated by the Cassini observation of ice dust grains with a salt admixture in the Saturnian E-ring (Postberg et al. 2009), Richterova et al. (2010a) used the last modification of the secondary emission model (Richterova et al. 2010b) to study the secondary emission yield and surface potential of icy dust grains as a function of the grain size and the salt admixture (see Fig. 11). They find that the potential of the salty grain differs only very slightly from that calculated for the grain from pure water ice for any reasonable salt admixture and suggest that the presence of the salt changes only the scattering of the primary electrons inside the grain but does not affect the emission of the true secondary electrons. Hence, the parameters for the pure ice can be used for estimation of the charging and the presence of the salt can be neglected. Experimental data on ice dust grains are still missing, though.

The discussed measurements were carried out at grains with small surface charge. The grain potential, however, influences SEE. Even more, a large negative electric field at the

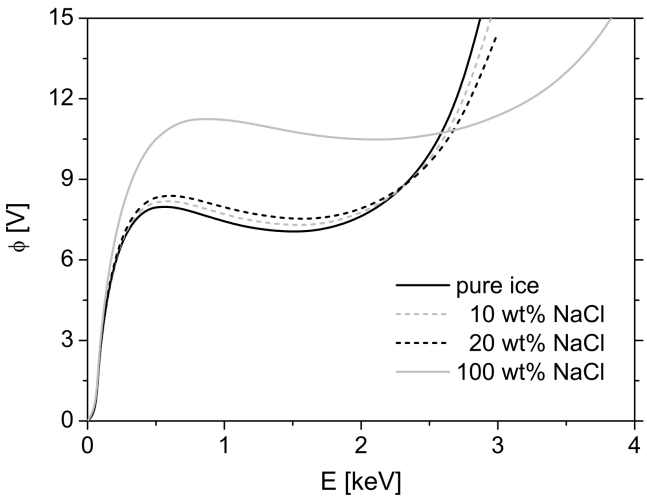


Fig. 11 The surface potentials, ϕ as a function of the primary electron beam energy for ice with different admixtures of NaCl. The admixture is given in percents of weight.

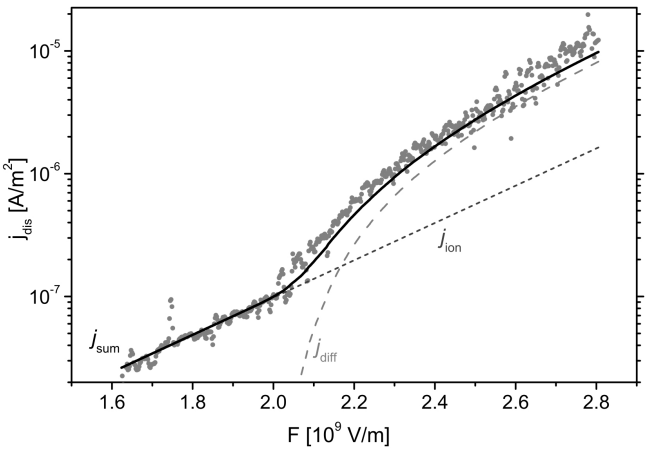


Fig. 12 Comparison of the measured discharging of the carbon grain with the simple model of FIE. The discharging current, j_{dis} is composed of two terms: j_{ion} and diffusion current, j_{dif} and their sum is denoted as j_{sum}. (Adopted from Jerab et al. (2010b). Reprinted with permission ©2010 IEEE.)

grain surface modifies electronic structure of the surface states and through increase of the escape probability of excited electrons change the SEE (Beranek et al. 2009).

5.2.4 Ion beam experiments

Another class of laboratory experiments was devoted to interaction of the dust with energetic ions. It is generally expected that ion beams would charge the dust grain to a potential equal to the beam energy but Cermak et al. (1995) have shown that a high grain potential can lead to spontaneous discharging, which is possibly attributed to field evaporation of bulk material. Sternovsky et al. (2001) found a threshold electric field of the order of 10^8 V/m

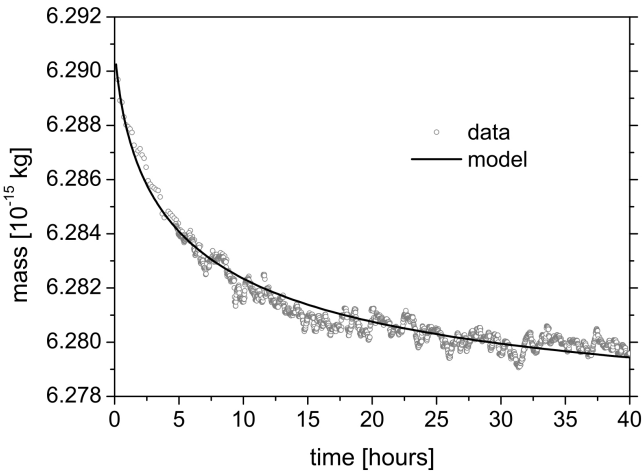


Fig. 13 The change of the grain mass in time. The parameters of the model ($D = 5 \cdot 10^{-16}$ cm$^2/s$, treatment: 8 h, incident current: 32 400 particles per second). The horizontal axis indicates the time after treatment switch-off. (Adopted from Beranek et al. (2010b). Reprinted with permission ©2010 IEEE.)

for the field evaporation current. The authors concluded that this process can lead to ion emission from submicron grains exposed to solar UV radiation.

New systematic experiments show that the threshold electric field as well as the discharging current depend on the charging history (Velyhan et al. 2004, Pavlu et al. 2006). The experiments used the melamine-formaldehyde spheres as dust grain samples and the structure of this material can be easily destroyed by the ion bombardment. For this reason, Jerab et al. (2007) repeated the charging experiment using gold samples. They found that the discharging current increases with the dose of ions applied to the grain. This effect was attributed to an implantation of the beam ions into the grain surface and their backward diffusion and consecutive field ionization of desorbed atoms. This conclusion was confirmed by further measurements that used carbon dust grains and Ar ions (Jerab et al. 2010). The authors developed a simple model of the grain discharging that takes into account the field ionization of the residual atmosphere in the vacuum vessel and diffusion and field ionization of the implanted ions. A good agreement between the model and experiment is demonstrated in Fig.12 where the discharging current is plotted as a function of the surface electric field (Note that actually both quantities change in time). The discharging current is supposed to be a sum of two components: the field ionization current, j_{ion} and diffusion current, j_{dif}.

Beranek et al. (2010b) performed a quantitative investigation of ion implantation. They used the combination of light carbon grains and heavy argon ions and measured directly the change in grain mass after prolonged ion bombardment. The results plotted in Fig. 13 show time constants of this process of the order of 10 hours at room temperature. Since the diffusion coefficients depend on the temperature exponentially, they concluded that the dust grains keep the implanted ions over long periods of time.

5.2.5 Sputtering experiments

Sputtering experiments are difficult and time consuming because the ion beam intensity should be kept low; otherwise the ion drag would blow the grain from the trap. Pavlu et al. (2007) bombarded a gold grain with Ar ions for 250 hours and this bombardment led to a

decrease of the grain mass by about one half. A model developed for an interpretation of the results revealed two factors that distinguish the sputtering yield of small grains from that measured on planar surfaces under the same conditions. First of them is the shape effect that increases sputtering yield by a factor of ≈ 2. The second factor is the grain charge resulting in the intense electric field at the grain surface. The electric field enhanced the sputtering yield by another factor of ≈ 2.5. Based on these results, Pavlu et al. (2008) computed the life-time of grains under typical conditions in the interplanetary space and found a value of $\approx 5.10^2$ years as a typical value for sputtering of a micron sized grain. The authors considered an average elemental composition of the solar wind and they have shown that heavy ions contribute to the overall sputtering yield by about one half due to their large efficiency. However, we should note that the real sputtering yield would be even higher because the influence of the implanted ions on the grain mass was neglected in the experiment and it can be significant. The sputtering rates derived from the laboratory measurement are higher than those discussed in section (4.1) and this is due to the sputtering from the heavy solar wind species. Another effect that was found and that was not taken into account in calculations is the ion drag effect. Energetic ions are often present as directional beams in the planetary magnetospheres or interplanetary space. The ion drag orients the grains along the longest axis into the beam direction and thus a large portion of the grain surface is bombarded by the ions with a small incident angle. Depending on the shape of a particular grain, this effect can bring a further increase of the sputtering rate by a factor of 2-5. The resulting dust lifetimes are still of the order of the Poyinting - Robertson lifetime, but the results shows that sputtering is more important than suggested by the theoretical considerations.

6 Dust in-situ measurements

Most of the instruments designed for in-situ detection of dust particles rely on either measurements of the dust primary surface charge or on measurements of the charge generated by the dust impact on a target. Knowledge of the dust charging and of the dust impact processes is therefore essential, and the limitations of available theory and experiment poses currently a limit to the data evaluation. Moreover, the Langmuir probe type dust measurements are in principle limited. Likewise the detection of dust with plasma wave measurements requires understanding the impact generated cloudlets and their evolution in the solar wind plasma. Optical detection methods are rarely applied and limited due to the small signal of scattered light in the case of the smaller dust, as well as by the time resolution of the light detection and by the requirements for deep space probes.

We discuss the dedicated dust instruments on sounding rockets in the ionosphere (6.1) and on deep space missions (6.2) and finally the dust detection with plasma experiments (6.3).

6.1 Dust instruments on sounding rockets

Dust can be observed in situ from sounding rockets in the ionosphere and this can in principle be made by mass spectrometry, like for the ions. Arnold et al. (1982) discuss the connection to a meteoric source in the context of measuring heavy ions ($m > 100$ amu) from sounding rocket. A possible detection of meteoric smoke particles from sounding rocket was reported by Schulte and Arnold (1992) based on quadrupole mass spectrometry measurements. They detected a layer of negative ions with masses up to 400 amu and above and

suggested that the ions ″are best explained as being the negatively charged fraction of the so-called meteor smoke particles″. Since then, dust measurements were made from rockets at altitudes roughly 70 to 100 km. The first measurements of dust particles with a dedicated instrument from sounding rocket was reported by Havnes et al. (1996). This instrument (as do later measurements e.g. Horanyi et al. 2000) detects the current due to the surface charge of impacting particles. The instrument contains two entrance grids that are biased to positive, respectively negative potential relative to the charge-collecting Faraday cup. The grid-induced field deflects charged species below a certain mass threshold to enter the detector and prevents the detection of electrons and light ions. Similarly Gelinas et al (1998) subtract the dust current from the larger background current with an oscillating electric field that alternately deflects the dust to two different anodes. The deflection only applies to a specific mass interval of particles. A similar concept with an applied alternating voltage was used to dected positively and negatively charged dust (Lynch et al. 2005). Also other deflection mechanisms were used. Some experiments combine the charge measurements with Langmuir probe measurements in order to better discriminate the ion and electron signals from the dust signals (Mitchell et al. 2001, Goldberg et al. 2001). Improved instruments applying primary charge detection are equipped with an UV flash lamp to enhance the dust charge by photoionisation and hence lower the threshold of detected dust particles (Rapp et al. 2003, 2010). Measurements with UV photometers of the solar light scattered at dust provide information on dust downward to 40-50 nm size and indicate the existence of smaller dust (Gumbel et al. 2001), but optical detection is limited by the light scattering efficiency of the dust particles at the wavelength of observation.

The dust instruments that measure primary charge detect, similarly to Langmuir probes the currents in a plasma, the combined current of ions, electrons and dust charge. They do not allow single particle detection and the dust masses are derived assuming a specific surface charge of the dust; hence the derived numbers are based on assumptions concerning the charging process. The data are possibly contaminated by secondary charges: The dust impact velocities range from several 100 m/s to 1 km/s and the dust impacts may generate free charges. This was already pointed out in the context of the first measurements (Havnes 1996) and Havnes and Næsheim (2007) re-evaluted the conditions of the measurements. For a different rocket experiment Barjatya and Swenson (2006) find that the data are influenced by triboelectric charging of the dust at the payload surface and they question in the reliability of Langmuir probe - type dust measurements in general.

Measurements from sounding rocket are moreover principally limited by the aerodynamics. Horanyi et al.(1999) find in a study for a specific experimental configuration that nano dust is deflected from entering the detector and that measurements of ice particles are further influenced by sublimation and condensation. Gumbel et al.(2001) point out that this may reduce the detection efficiency of the nano dust with sizes <10 nm by 10 to 100% depending on the individual detection geometry and atmospheric conditions and that further interactions within the instrument may influence the signal. They suggest combined dust flux and aerodynamic studies to attack this problem. Considering the aerodynamics of the meteoric smoke is further complicated by the fact that the payload moves from continuum flow to free molecular flow conditions (Hedin et al. 2007a).

Collecting dust samples would allow to prove the origin of the dust and this is the purpose of a rocket campaign project (Gumbel et al. 2005). A sampling device was designed to collect dust from a sounding rocket under minimized aerodynamical perturbations (Hedin et al. 2007b), but this failed so far for various technical reasons.

6.2 Dust instruments on deep space spacecraft

Dust in the interplanetary medium is predominantly observed with impact ionization detectors (for the technical and functional principles see Auer 2001, for a review of current and future dust in-situ measurements see Grün et al. 2005). The particles are destroyed during the impact and the produced electrons and ions that originate from the dust and from the detector target are measured. For each dust impact a sequence of different charge signals is measured within the instrument and allows to distinguish dust signals from those of other particles that possibly reach the instrument. The charges are typically measured at the entrance grids, the target and the sidewalls. Charge amplitudes provide information about the ionized mass fraction and the time sequence of the signals about the impact speed, if combined with pre-flight laboratory calibration measurements. Combination of the impact speed with the detector geometry and spacecraft velocity and orientation at the time of the event allow estimating the dust velocities.

Detailed numerical studies of the impact ionization process were performed (Hornung et al. 2000 and references therein). Laboratory measurements of impact generated charges are carried out during instrument calibration. Numerical and experimental results agree with the following rough picture: At impact velocities 1 – 50 km/s, typical in the interplanetary medium, collisions lead to fragmentation as well as to the partial ionization of dust and target surface materials. At impact velocities beyond 50 km/s the entire particle is destroyed and the degree of ionization of the produced vapor increases with impact velocity. For large impact velocities and small particles the detectors provide reliable information about dust mass and in some cases relative velocities. Major limitation of the measurements lie in the complexity of the impact ionization process and in the range of dust impact velocities, which do not extend above 100 km/s in laboratory measurements. Moreover, the dust masses, materials and structures that can be prepared for laboratory calibration measurements are limited (see also section 6.3).

For recent measurements in the interplanetary medium(see Krüger et al. 2006 and references there) the mass range of particles for which the mass could be derived is 10^{-19}–10^{-10} kg and the determined velocity has a factor of 2 uncertainty and the mass a factor of 10 uncertainty. Note, that a significant fraction of registered events is caused by impacts of fast and small dust particles outside of the range of instrument calibration.

The element composition of the dust can be inferred in principle from impact ionization experiments measuring the time of flight of the produced ions. The detected species originate from complex ionization, expansion and recombination processes and this complicates the analysis. So far a full quantitative analysis of dust composition in space has not been achieved. The mass spectra generated by dust impacts were measured and analyzed during the space missions to comet Halley and also in the interplanetary medium with the Cosmic Dust Analyser (CDA) experiment onboard the Cassini spacecraft (see Hillier et al. 2007; Postberg et al. 2008).

In-situ measurement of dust surface charges in the interplanetary medium were firstly achieved also with CDA (Kempf et al. 2004). The dust passes a system of entrance grids while entering CDA and thereby induces a charge signal. In the case of 6 reported impact events the dust charge was sufficiently large to generate a clear feature of induced charge at the grid system and this allowed to derive the surface charge. The measured surface charges range between $1.3 - 5.4 \cdot 10^{-15}$ C. The mass of the particles is uncertain. The authors suggested their results agree with theoretical models of dust surface charging: assuming a surface potential of + 5 V for the dust, they find that the detected dust particles have a mass

larger than 10^{-13} kg and that the number of detected particles agrees with model predictions in this mass range.

6.3 Dust detection with plasma experiments

Dust detection by plasma instruments is based on measuring the voltage induced on an electric antenna when a dust particle impacts a spacecraft at high speed and then vaporises and ionizes, as well as a part of the targets material. This produces a plasma cloud whose electric charges are detected. The released charge which is much larger than that normally carried by the dust particle induces a large voltage on the electric antenna (Fig. 14).

Such a detection was made serendipitously for the first time on the spacecraft Voyager 1 in Saturn's E ring (Aubier et al. 1983; Gurnett et al. 1983), and was later used in various environments of the heliosphere (see Meyer–Vernet 2001 and references therein).

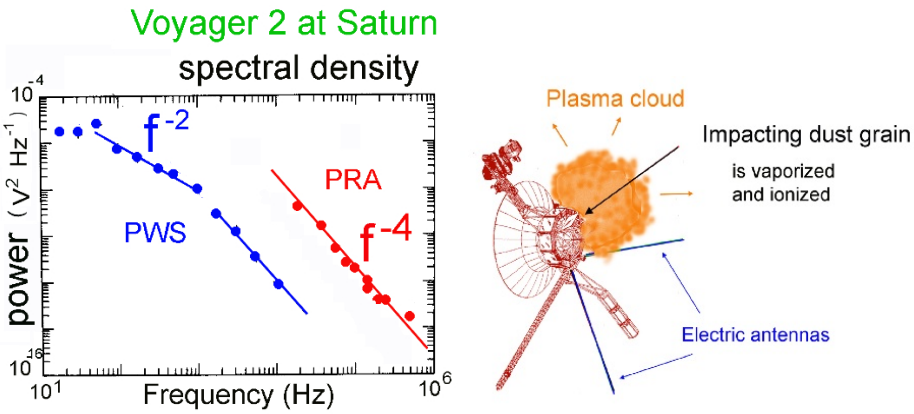


Fig. 14 A dust particle impacting a spacecraft at a high speed vaporises and ionizes; this produces an expanding plasma cloud whose charge is partially recollected, inducing a transient voltage on the electric antennas. The voltage spectral density measured on Voyager 2 by the radio (denoted as PRA) and plasma wave instruments (denoted as PWS). The spectrum varies as f^{-4} at high frequencies and f^{-2} at lower frequencies. Because of the asymmetrical antenna configuration, the amplitude is much greater when the antennas are in monopole mode (the case of the Voyager PRA instrument).

This process is basically the same as for classical impact ionization detectors (Section 6.2) with, however, two important differences. Firstly, since the wave instruments on present space probes were generally not designed to measure dust, they were not calibrated for that purpose, so that the dust measurement is generally not accurate - a problem which may be solved on future space probes. However, this technique has the advantage of having a much larger effective area than conventional dust detectors, since it is of the order of magnitude of the cross-section of the spacecraft itself, which generally amounts to several squared meters. The charge released upon impact increases fast with the impact speed (see Section 4.1). This produces voltage pulses on the antenna, since charges are recollected by the spacecraft or antennas and an electric field is produced. The voltage V measured as a function of Q depends on the antenna/spacecraft configuration and on the impact site; it rises with a time scale τ_r, which can be evaluated from simple physical arguments (Meyer–Vernet 1986; Oberc 1994, 1996), and subsequently decays with a longer time scale τ_d determined

by the system time constant. Measuring these impulses with a wave-form receiver reveals the mass and impact rate of the particles. At frequencies $f \gg 1/2\pi\tau_r$, the Fourier transform $|V(\omega)|$ is determined by the discontinuity of the derivative in the rising part, so that it is given by

$$|V(\omega)| = V_{\max}/(\tau_r \omega^2) \tag{20}$$

Hence, the total voltage power spectral density V_ω^2 at high frequencies is (Meyer–Vernet 1996)

$$V_\omega^2 = 2\langle N V_{\max}^2/\tau_r^2\rangle/\omega^4 \tag{21}$$

where N is the impact rate, the brackets stand for an average over the dust mass distribution, and the factor of two stems from the fact that the spectral density is defined for positive frequencies only. This yields a power spectrum varying as $1/f^4$ at high frequencies, whose level can be used to deduce the dust mass distribution. At frequencies intermediate between the inverse decay time and the inverse rise time, the signal behaves as a step function, producing a $1/f^2$ power spectrum, which is not adequate for detecting dust because it may be confused with genuine plasma signals having a similar spectral index (Meyer–Vernet and Perche 1989).

The corresponding spectral density is shown on Fig. 14, as recorded with the radio (denoted as PRA; Aubier et al. 1983; Meyer–Vernet et al. 1998) and plasma wave (denoted as PWS; Gurnett et al. 1983; Tsintikidis et al. 1994) instruments on Voyager 2. When the antenna is operated as a dipole, the receiver records the difference of potential between the two antenna arms. Hence, except when the antenna is short and/or the arms are not symmetrically located with respect to the spacecraft, it responds weakly to dust impacts on the spacecraft and this response depends on the asymmetries of the system. This is illustrated on Fig. 14, where the signal of the plasma wave instrument (PWS, operating as a dipole) is much smaller than the one recorded by the radio instrument (PRA, operating as a monopole, i.e. measuring the difference of potential between an antenna arm and the spacecraft).

This technique has been recently applied to the measurement of nanoparticles in the solar wind near Jupiter with the wave instrument (RPWS) on the spacecraft Cassini (Meyer-Vernet et al. 2009a) and at 1 AU with the WAVES instrument on STEREO (Meyer-Vernet et al. 2009b)

Two other techniques have been proposed to detect dust with plasma instruments, which do not require the dust particles to impact the spacecraft. They are based on the transient voltage that the charge carried by the particle induces on an electric antenna when it passes closer to it than the ambient Debye length. The dust particles can be detected individually by a wave-form analysis (Meuris et al. 1996). The main advantage of this latter technique is that it has an extremely large equivalent area for dust detection - roughly the product of the antenna antenna length by the Debye length, which amounts typically to 100 m^2 in the interplanetary medium. Since, however, the voltage induced is small and might be confused with genuine plasma wave effects, this method requires a very sensitive receiver and an elaborate analysis. When the dust concentration is sufficiently high that numerous flybys are detected simultaneously, the wave receiver records a voltage power spectrum which is a superposition of the spectra produced by individual particle flybys near the antenna (Meyer–Vernet 2001). The principle is similar to the electron measurements performed routinely from plasma thermal noise spectroscopy in space (Meyer–Vernet and Perche 1989), but the calculation is easier because the dust particles generally move so slowly that the plasma temporal dispersion plays a negligible role.

7 Summary and discussion:

In spite of meteors well visible in the night sky, at present we do not know very well the amount and composition of meteoroid species that are fed into the atmosphere. The observation of radar meteors in general is an important tool to better quantify the influx of meteoroid mass into the Earth atmosphere and its spatial and temporal variations. We can roughly estimate the overall element composition of the entering meteoroids from cosmic abundance considerations and laboratory analyses of collected samples of dust and meteorites provide information about the solids that survive the entry process. Meteor spectra provide further information, but do not reveal the full meteoroid composition.

The perspectives of future observations of dusty plasma effects are diverse. Theory is still insufficient to explain the observed phenomena and knowledge from dust experiments and from laboratory data is limited. Dust charging and dust impact processes play an important role in the dust plasma interactions, but also for the understanding of dust in-situ instruments. In both cases there is a gap between the theory and the laboratory experiments and furthermore the laboratory experiments do not fully reproduce space conditions.

Dusty plasma effects comprise (i) the alteration of the dust as a result of the ambient plasma (charging, erosion, condensation and growth), (ii) the generation of neutrals, electrons or ions from dust, (iii) the change in energy or charge state of plasma species by interaction with the dust, (iv) the dust motion or change of motion as a result of the ambient plasma, and (v) collective behavior caused by the presence of dust. The latter is often denoted as dusty plasma in the narrow sense of the expression. We summarize our review according to these listed dusty plasma effects (i - v) and will point out options for future space experiments and theoretical and experimental studies.

Referring to (i), the alteration of dust in the Earth's ionosphere in the case of meteoroid ablation is inferred from the observed meteor phenomenon. The surface charge of dust is (in a few cases) detected with in-situ dust measurements (both in interplanetary space and Earth's ionosphere). At present, however, the measurements are subject to a large uncertainty. We are not aware of direct measurements of dust erosion, dust condensation and dust growth in natural cosmic environments (aside from experiments under microgravity conditions, which are beyond the scope of this review). A major limitation is the paucity of laboratory data on the interactions of solids (such as secondary electron emission yields, sputtering yields) and limited knowledge of the small particle effects that may influence the interaction. Direct measurements of the dust surface charge will be most likely possible with improved space dust in-situ measurements.

The generation of neutrals, electrons or ions from dust (ii) is possibly detected in the interplanetary medium via the solar wind pick up ions. The vapor production by mutual dust collisions is a major mechanism. The generation of neutrals, electrons or ions from dust leads to the different meteor phenomena in the Earth's ionosphere. Experimental techniques have recently improved enough to carry out detailed spectroscopic meteor observations, which are a promising tool to shed some light on the nature of the generated species. Such studies may also lead to information about the composition and, perhaps, the shapes of meteoroids. We point out that a consistent physical model to describe meteor spectra is still missing, though.

The change in energy or charge state of plasma species by interaction with the dust (iii) is suggested (aside from the release of dust species) in order to explain pick up ions and also neutral solar wind, but so far the observed species could not be attributed to a dust related process in a quantitative way. Also in the case of high altitude meteors such processes play a role. Future experimental studies that may lead to some progress are the observational

studies of high altitude meteors, as well as the simultaneous in- situ detection of dust and plasma components in the interplanetary medium.

Referring to (iv), the dust submicrometer-sized dust is significantly influenced by radiation pressure and gravitational forces. Fast nano dust particles accelerated by solar wind interactions have also been detected. Though the forces acting on dust in a plasma are well known, the uncertainty in the parameters that determine these forces hampers a quantitative discussion.

Referring to (v), many theoretical studies consider plasma collective effects in the Earth's ionosphere and enhanced dust densities in meteor trails potentially may lead to waves and instabilities. The development of quantitative models is limited by the lack of experimental data concerning the detection of dust in the nm size range. As a result of the small dust number densities compared to the ambient solar wind plasma densities plasma collective effects are not likely to occur in the interplanetary medium on a large scale, though they may occur locally, as for instance in the vicinity of comets. The collective effects in meteor trails may in the future be possibly detected by radar observations.

In summary we identify the following issues for future research on dust plasma interactions: (a) ion, atom, and molecule interactions on dust surfaces in the limit of low energies (sputtering, recombination, molecular reactions); (b) the charging mechanisms of nano-dust; (c) the dynamics of nano-dust including the influence of charge fluctuations; (d) the impact ionization process of nano-dust; (e) the sublimation of nano-dust. These research issues require theoretical considerations combined with laboratory measurements. The study of high altitude meteors may further reveal dust plasma interactions as well as dusty plasma collective effects, though as a result of the number of contributing parameters (like dust size, composition, charging processes) these seem to be more complex than predicted by current theory.

Progress in space measurements is expected from the in-situ detection of nano particles and of dust surface charges in the ionosphere and interplanetary medium, the detection of plasma waves related to charged dust or dust impacts and the detection of ion species generated by dust impact in the Earth's ionosphere and in the interplanetary medium.

Acknowledgements This article was prepared within an International Team at the International Space Science Institute, ISSI, Bern, Switzerland. Parts of this work were supported by the German Aerospace Center DLR (Deutsches Zentrum für Luft- und Raumfahrt) under project (RD-RX / 50 QP 0403) and by the Japanese Ministry of Education, Science, Sports and Culture, Grant-in-Aid for Scientific Research -C 19540475, 2008. This collaboration was also partly supported by the Swedish Foundation for International Cooperation in Research and Higher Education, STINT (I.M. and E.M.) and by the US DOE Grant No. DE-FG02-04ER54804 and NSF Grant no. PHY-0903808 NSF and ATM-0907941 (MR) and by the Open Research Center at Kindai University. We acknowledge helpful discussions with Markus Rapp in the initial stage of this project and we thank Dennis Höning for his support in preparing the manuscript.

References

1. M. Abbas, P. Craven, J. Spann, E. West, J. Pratico, D. Tankosic, and C. Venturini, Phys. Scr. T98, 99-103 (2001).
2. M. Abbas, P. Craven, J. Spann, D. Tankosic, A. LeClair, D. Gallagher, E. West, J. Weingartner, W. Witherow, and A. Tielens, Astrophys. J. 614(2), 781-795 (2004).
3. M. M. Abbas, D. Tankosic, P. D. Craven, J. F. Spann, A. LeClair, E. A. West, J. C. Weingartner, A. G. G. M. Tielens, J. A. Nuth, R. P. Camata, and P. A. Gerakines, Astrophys. J. 645 (1), 324-336 (2006).
4. M. M. Abbas, D. Tankosic, J. F. Spann, M. J. Dube, and J. A Gaskin, Space Technology and Applications International Forum-STAIF 2008, AIP, CP969, 2008.
5. C.W. Allen, Astrophysical Quantities, 3rd Edn. (London, Athlone, 1976)

6. Arnold et al. (1982)
7. M.G. Aubier, N. Meyer–Vernet, B.M. Pedersen, Geophys. Res. Lett. 10, 5 (1983)
8. S. Auer, in Astron. Astrophys. Library, ed. By E. Grün, B.A.S. Gustafson, S. Dermott, H. Fechtig (Springer, Berlin and New York, 2001) pp. 385–444
9. S. Auer, in Workshop on Dust in Planetary Systems. September 26-30 2005, Kauai, Hawaii. ed. By H. Krüger and A. Graps (ESA SP, 2007) p. 231-233
10. P.B. Babadzhanov, Astron. Astrophys. 384, 317-321 (2002)
11. A. Barjatya and C.M. Swenson, J. Geophys. Res. 111, A10302, doi:10.1029/2006JA011806 (2006)
12. P.M. Banks, J. Geophys. Res. 19, 4341 (1971)
13. A. Barkan, N. D'Angelo, and R. Merlino, Phys. Rev. Lett. 73, 3093-3096 (1994).
14. R. Behrisch and W. Eckstein, in Topics Appl. Physics 110, ed. By R. Behrisch, W. Eckstein (Springer, Berlin Heidelberg, 2007)
15. L.R. Bellot Rubio, M.J. Martninez Gonzalez, L.R. Herrera, J. Licandro, D.M. Delgado, P.R. Gil, M. Serra-Ricart, Astron. Astrophys. 389, 680-691(2002)
16. M. Beranek, I. Richterova, Z. Nemecek, J. Pavlu, and J. Safrankova, Eur. Phys. J. D, 54 (2), 299-304 (2009).
17. M. Beranek, M. Vysinka, J. Pavlu, I. Richterova, Z. Nemecek, J. Safrankova, IEEE Trans. Plasma Sci. 38 (4), 886-891 (2010a).
18. M. Beranek, J. Pavlu, J. Vaverka, I. Richterova, J. Safrankova, and Z. Nemecek, J. Geophys. Res., in press, 2010b.
19. P.A. Bernhardt, G. Ganguli, M.C. Kelley, W.E. Swartz, J. Geophys. Res. 100, 23811 (1995)
20. D. Bilitza and B.Reinisch, J. Adv. Space Res. 42, 599-609 (2007) doi:10.1016/j.asr.2007.07.048
21. J. Borovička, Astron. Astrophys. 279, 627-645 (1993)
22. J. Borovička, in Meteorids 1998, ed. By W.J. Baggaley and V. Porubcan (Astron. Inst., Slovak Acad. Sci., 1999) pp. 355-363
23. Borovicka, J., Roya. J.: Astron. Soc. Canada 100 (2006) 194198.
24. M. Bzowski and M. Krlikowska, in: (B. Fleck et al. eds.) Proceedings of the Solar Wind 11 / SOHO 16 Conference, ESA SP-592) 397-400 (2005)
25. U. Brändström, The Auroral Large Imaging System - Design, Operation, and Scientific Results, IRF Sci. Rep. 279, Swedish Institute of Space Physics, Kiruna (2003)
26. V.A. Bronshten, Physics of Meteoric Phenomena, (D. Reidel. Publ.Co., Dordrecht, 1983), p. 356
27. N. Brosch, L.S. Schijvarg, M. Podolak, M.R. Rosenkrantz, in Proceedings of the Meteoroids 2001 Conference, 6-10 August 2001, Kiruna, Sweden, ed. By B. Warmbein, (ESA Publications Division, Noordwijk, 2001), pp. 165-173
28. N. Brosch, D. Polishook, R. Helled, S. Schijvarg, M. Rosenkrantz, Atmos. Chem. Phys. 4, 1063-1069 (2004)
29. P.G Brown, M.D. Campbell, R.L Hawkes, C. Treijsmeijer, J. Jones, Planet. Space Sci. 50, 45-55 (2002)
30. M. D. Campbell-Brown, Earth, Moon, and Planets, 95, 521-531 (2004)
31. M.D. Campbell-Brown and D. Koschny, Astron. Astrophys. 418, 751-758 (2004)
32. M.D. Campbell-Brown and S. Close, Monthly Notices of the Royal Astronomical Society 382, 1309-1316 (2007)
33. M.D Campbell, P.G. Brown, A.G. LeBlanc, R.L. Hawkes, J. Jones, S.P. Worden, R.R. Correll, Meteoritics & Planetary Science 35, 1259-1267 (2000)
34. J.D. Carpenter, T.J. Stevenson, G.W. Fraser, J.S. Lapington, D. Brandt, J. Geophys. Res. 110 (2005)
35. Z. Ceplecha, J. Borovička, W. G. Elford, D. O. Revelle, R. L. Hawkes, V. Porubcan, M. Simek, Space Sci. Rev. 84, 327-471 (1998)
36. I. Cermak, E. Grün, and J. Svestka, Adv. Sp. Res. 15(10), 59-64 (1995).
37. E. Chapin and E. Kudeki, Geophys. Res. Lett. 21, 2433 (1994)
38. J.Y.N. Cho and J. Röttger, J. Geophys. Res. 102, 2001-2010 (1997)
39. V.W. Chow and M. Rosenberg, Phys. Plasmas 3, 1202 (1996)
40. V. Chow, D. Mendis, and M. Rosenberg, J. Geophys. Res. 98, 19065 (1993).
41. V. Chow, D. Mendis, and M. Rosenberg, IEEE Trans. Plasma Sci. 22, 179 (1994).
42. V.W. Chow, et al., Astrophys. Space Sci. 223, 144 (1995)
43. S. Close, S.M. Hunt, M.J. Minardi, F.M. McKeen, Radio Science 35, 1233-1240 (2000) bibitemCollier2003 M.R. Collier, T.E. Moore, K. Ogilvie, D.J. Chornay, J. Keller, S. Fuselier, J. Quinn, P. Wurz, M. Wüest, K.C. Hsieh, AIP Conference Proceedings, 679, 790-793 (2003)
44. S.G. Coulson, Monthly Notices of the Royal Astronomical Society 332, 741-744 (2002)
45. S.G. Coulson, N.C. Wickramasinghe, Monthly Notices of the Royal Astronomical Society 343, 1123-1130 (2003)
46. C.L. Croskey, J.D. Mitchell, M. Friedrich, K.M. Torkar, U.P. Hoppe, R.A. Goldberg, Geophys. Res. Lett. 28, 1427-1430 (2001)

47. A. Czechowski and I. Mann, Astrophys. J. 714, 89-99 (2010)
48. P. Czechowsky, R. Ruester, G. Schmidt, Geophys. Res. Lett. 6, 459-462 (1979)
49. D.J. Cziczo, D.S. Thomson, D.M. Murphy, Science 291, 1772-1775 (2001)
50. N. D'Angelo, Planet. Space Sci. 38, 1143 (1990)
51. N. D'Angelo, Planet. Space Sci. 51, 393 (2003)
52. N. D'Angelo, Phys. Lett. A 336, 204 (2005)
53. J.S. Dohnanyi, in: Cosmic Dust (J. A. M. McDonnell, Ed.), pp. 527-606. Wiley, Chichester (1978)
54. B. Draine and E. Salpeter, Astrophys. J. 231(1), 77-94 (1979).
55. B.T. Draine and B. Sutin, Astrophys. J. 320, 803 (1987)
56. J. Duprat, C. Engrand, M. Maurette, G. Kurat, M. Gounelle, C. Hammer, Adv. Space Res. 39 (4), 605-611(2007).
E. Dubinin, M. Fraenz, J. Woch, J. Winnigham, R. Frahm, R. Lundin, and S. Barabash, Planet. Space. Sci. 56 (6), 846-851 (2008).
57. L.P. Dyrud, M.M. Oppenheim, A.F. vom Endt, Geophys. Res. Lett. 29, 2012 (2002)
58. L. Dyrud, D. Wilson, S. Boerve et al. Adv. Space Res. 42, 136-142 (2008)
59. H.J. Fahr, H.W. Ripken, G. Lay, Astron. Astrophys. 102, 359 (1981)
60. V.E. Fortov, A.V. Ivlev, S.A. Khrapak, A.G. Khrapak, G.E. Morfill, Phys. Reports 421, 1 (2005)
61. Y. Fujiwara, M. Ueda, Y. Shiba, M. Sugimoto, M. Kinoshita, C. Shimoda, Geophys. Res. Lett. 25, 285-288 (1998)
62. T. Fukuzawa, S. Hashim, Y. Matsuoka, et al., J. Appl. Phys. 86, 3543 (1999)
63. P. Gabrielli, C. Barbante, J. Plane et al., Nature 432, 1011-1014 (2004)
64. L.J. Gelinas, K.A. Lynch, M.C. Kelley, S. Collins, S. Baker, Q. Zhou, J.S. Friedman, Geophys. Res. Lett. 25, 4047-4050 (1998)
65. L.J. Gelinas, K.A. Lynch, M.C. Kelley, R.L. Collins, M. Widholm, E. MacDonald, J. Ulwick, P. Mace, J. Geophys. Res. 110 (2005)
66. G. Gloeckler, L.A. Fisk and J. Geiss, in: (M. Maksimovic, K. Issautier, N. Meyer-Vernet, M. Moncuquet, F. Pantelliniet Eds.) AIP CP 1216, 514-517 (2010)
67. C.K. Goertz, Rev. Geophys. 27, 271 (1989)
68. R.A. Goldberg, R.F. Pfaff, R.H. Holzworth, et al. Geophys. Res. Letters 28 no. 8, 1407-1410 (2000)
69. R.J.L. Grard, J. Geophys. Res. 78, 2885 (1973)
70. E. Grün, H.A. Zook, H. Fechtig, R.H. Giese, Icarus 62, 244 (1985)
71. E. Grün, R. Srama, H. Krüger, et al., Icarus 174, 1 (2005)
72. J. Gumbel, Journ. Geophys. Res. 106 no A6, 10553-10563 (2001)
73. J. Gumbel, K.W.T. Waldemarsson, F. Giovane, et al., Proc. 17th ESA Symposium on European Rocket and Balloon Programmes and Related Research (ESA SP-590), 139-144 (2005)
74. D.A. Gurnett, E. Grün, D. Gallagher, et al., Icarus 53, 236 (1983)
75. O. Havnes, Astron. Astrophys. 193, 309 (1988)
76. O. Havnes and F. Sigernes, Journal of Atmospheric and Solar-Terrestrial Physics 67, 659-664 (2005)
77. O. Havnes and L.I. Næsheim, Ann. Geophys., 25, 623637 (2007)
78. O. Havnes, T. Aslaksen, A. Brattli, Phys. Scripta T89 133 (2001)
79. O. Havnes, J. Trøim, T. Blix, W. Mortensen, L.I. Næsheim, E. Thrane, T. Tønnesen, J. Geophys. Res. 101, 10839-10848 (1996)
80. R.L. Hawkes and J. Jones, Monthly Notices of the Royal Astronomical Society 173, 339-356 (1975)
81. R.L. Hawkes, in Meteors in the Earth' s Atmosphere, ed. By E. Murad, E. and I. P. Williams (Cambridge University Press, 2002), pp. 97-122
82. R.L. Hawkes, P.G. Brown, N.R. Kaiser, A.J. Faloon, K.A. Hill, L.A. Rogers, EMP 95, 587-593 (2004)
83. A.E. Hedin, J. Geophys. Res. 96, 1159, (1991)
84. J. Hedin, J. Gumbel, and M. Rapp, Atmos. Chem. and Physics 7, 3701-3711 (2007)
85. J. Hedin, J. Gumbel, T. Waldemarsson, and F. Giovane, Adv. Space Res. 40, 818-824 (2007)
86. J.R. Hill and D.A. Mendis, Canad. J. Phys. 59, 897 (1981)
87. K.A. Hill, L.A. Rogers, R.L. Hawkes, Earth, Moon, and Planets 95, 403-412 (2005)
88. J.K. Hillier, S.F. Green, N. McBride et al., Icarus 190, 643-654 (2007)
89. R.H. Holzworth, R.F. Pfaff, R.A. Goldberg, et al., Geophys. Res. Lett. 28, 1435 (2001)
90. M. Horanyi and D.A. Mendis, Astrophys. J. 294, 357-368 (1985)
91. M.S. Horanyi, B. Robertson, J. Smiley, G. Gumbel, G. Witt, B. Walch, Geophys. Res. Lett. 27, 3825-3828 (2000)
92. M.S. Horanyi, Annu. Rev. Astron. Astrophys. 34, 383418 (1996)
93. K. Hornung, Y.G Malama, K.S Kestenboim, Astrophys. Space Sci. 274, 355-363 (2000)
94. D.W. Hughes, in Meteors and Cosmic Dust, ed. By J. A. M. McDonnell (Wiley-Interscience, Hoboken, N. J, 1978), pp. 123-186
95. D.M. Hunten, R.P. Turco, O.B. Toon, J. Atmos. Sci. 37, 1342-1357 (1980)

96. M. Jerab, I. Richterova, J. Pavlu, J. Safrankova, and Z. Nemecek, IEEE Trans. Plasma Sci. 35 (2), 292-296 (2007).
97. M. Jerab, J. Vaverka, M. Vysinka, Z. Nemecek, J. Safrankova, IEEE Trans. Plasma Sci. 38 (4), 798-802 (2010).
98. J. Jones and A.R. Webster, Planet. Space Sci. 39, 873-878 (1991)
99. A.P. Jones, A. G. G. M. Tielens, and D. J. Hollenbach, D. J., Astrophys. J. 469, 740-764 (1996).
100. G .H. Jones, A. Balogh, D. J. McComas, and R. J. MacDowall, Icarus 166, 297-310 (2003).
101. A. G. G. M. Tielens, C.F. McKee, C.G, Seab, and D. J. Hollenbach, D. J. Astrophys. J., 431, 321-340(1994).
102. O. Kalashnikova, M. Horanyi, G.E. Thomas, et al., Geophys. Res. Lett. 27, 3293 (2000)
103. T.J. Kane and C.S. Gardner, Science 259, 1297-1300 (1993)
104. M.C. Kelley, The Earth' s Ionosphere: Plasma Physics and Electrodynamics, (Academic Press, San Diego, 1989)
105. M.C. Kelley, C. Alcala, J.Y.N. Cho, Journal of Atmospheric and Solar-Terrestrial Physics 60, 359-369 (1998)
106. M.C. Kelley, Radio Sci. 39, RS2015 (2004)
107. S. Kempf, R. Srama, N. Altobelli, et al., Icarus 171, 317 (2004)
108. J. Kero, Swedish Institute of Space Physics, PhD theses, Kiruna, Sweden (2008)
109. J. Kero, C. Szasz, G. Wannberg, A. Pellinen–Wannberg, A. Westman, Geophys. Res. Lett. 35 (2008a) doi:10.1029/2008GL033402
110. J. Kero, C. Szasz, A. Pellinen–Wannberg, G. Wannberg, A. Westman, D. D. Meisel, Geophys. Res. Lett. 35 (2008b) doi:10.1029/2007GL0327333
111. J. Kero, C. Szasz, A. Pellinen–Wannberg, G. Wannberg, A. Westman, "3D position determination of small coherent meteor targets with EISCAT UHF," accepted for publication in Annales Geophysicae, 2008c
112. H. Kimura, I. Mann, E.K. Jessberger, Astrophys. J. 582, 846-858 (2003)
113. P. Koten, J. Borovika, P. Spurný, H. Betlem, S. Evans, Astron. Astrophys. 428, 683-690 (2004)
114. P. Koten, P. Spurný, J. Borovička , S. Evans, A. Elliott, H. Betlem, R. Štork, K. Jobse, Meteoritics & Planetary Science 41 , 1305-1320 (2006)
115. A.R. Krauss and D.M. Grün, J. Nuclear Materials 85-86, 1179-1183 (1979)
116. H. Krüger, N. Altobelli, B. Anweiler, S.F. Dermott, V. Dikarev, A. Graps, E. Grün, B.A. Gustafson, D.P. Hamilton, M.S. Hanner, M. Horanyi, J. Kissel, M. Landgraf, B.A. Lindblad, G. Linkert, I. Mann, J.A.M. McDonnell, G. Morfill, C. Polanskey, G. Schwehm, R. Srama, H.A. Zook, Planet. Space Sci. 54, 932-956 (2006)
117. A. Kull, E. Kopp, C. Granier, G. Brasseaur, J. Geophys. Res. 102, 9705-9716 (1997)
118. J.G. Laframboise and L. W.Parker, Phys. Fluids, 16, 629 (1973)
119. Ch, Leinert, S. Bowyer, L.K. Haikala, et al. Astron. Astrophys. Suppl. 127, 1-99 (1998)
120. J. Lemaire, Astrophys. J. 360, 288 (1990)
121. B.A. Lindblad, L. Neslušan, V. Porubčan, J. Svoreň, Earth, Moon, and Planets 93, 249-260 (2003)
122. S.G. Love and D.E. Brownlee, Science 262, 550-553 (1993)
123. K.A. Lynch et al., J. Geophys. Res. 110 (2005) A03302 doi: 10.1029/2004JA010502
124. I. Mann, in Landolt Boernstein Handbook Series, LB VI/4B: Chap. 4.3.2., ed. By J. Trümper (Springer, Heidelberg, 2009)
125. I. Mann and A. Czechowski, Astrophys. J. 621, L73-L76 (2005)
126. I. Mann, E. Murad, A. Czechowski, System, Planet. Space Sci. 55, 1000-1009 (2007)
127. I. Mann, H. Kimura, D. A. Biesecker, B. T. Tsurutani, E. Grün, B. McKibben, J. C. Liou, R. M. MacQueen, T. Mukai, L. Guhathakurta, P. Lamy, Space Sci. Rev. 110, 269 (2004)
128. I. Mann, A. Czechowski, and N. Meyer-Vernet, N. in: (M. Maksimovic, K. Issautier, N. Meyer-Vernet, M. Moncuquet, F. Pantelliniet Eds.) AIP CP 1216, 491-496 (2010a)
129. I. Mann, A. Czechowski, and N. Meyer-Vernet, A. Zaslavsky, H. Lamy H., Plasma Phys. Control. Fusion 52, in print (2010b).
130. G.G. Mateshvili, Y.D. Mateshvili, N.Y. Mateshvili, Sol. Syst. Res. 31, 483-488 (1997)
131. N. Mateshvili, G. Mateshvili, I. Mateshvili, L. Gheondjian, O. Avsajanishvili, Meteoritics & Planetary Science 34, 969-973 (1999)
132. J.D. Mathews, D. Janches, D.D. Meisel, Q.-H. Zhou, Geophys. Res. Lett. 28 , 1929-1932 (2001)
133. N. McBride and J.A.M. McDonnell, Planet. Space Sci. 47, 1005 (1999)
134. D.W.R. McKinley, Meteor Science and Engineering, (McGraw-Hill, New York, 1961)
135. W.J. McNeil, S.T. Lai, E. Murad, J. Geophys. Res. 103, 10899 - 10911 (1998)
136. W.J. McNeil, E. Murad, J.M.C. Plane, in Meteors in the Earth' s Atmosphere, ed. By E. Murad and I. P. Williams (Cambridge University Press, 2002), pp. 265-287
137. L. Megner, D.E. Siskind, M. Rapp, J. Gumbel, J. Geophys. Res. 113 D03202, doi:10.1029/2007JD009054 (2008)

138. L. Megner, M. Rapp, J. Gumbel, Atmos. Chem. Phys. 6, 4415 (2006)
139. D.A. Mendis and M. Rosenberg, Annu. Rev. Astron. Astrophys. 32, 419 (1994)
140. P. Meuris, N. Meyer-Vernet, J. F. Lemaire, J. Geophys. Res. 101, 24471-24478 (1996)
141. R.A. Mewaldt, C.M.S. Cohen, G.M. Mason, et al. Space Sci. Rev. 130, 207-219 (2007)
142. N. Meyer-Vernet, Astron. Astrophys. 105, 98 (1982)
143. N. Meyer-Vernet, Adv. Space Res. 5, 37 (1986)
144. N. Meyer-Vernet, in Proc. 7th Spacecraft Charging Technology Conf. ed. By R. A. Harris, ESA SP-476 (2001), pp. 635-640
145. N. Meyer-Vernet, Basics of the Solar Wind, Cambridge University Press (2007)
146. N. Meyer-Vernet and C. Perche, J. Geophys. Res. 94, 2405 (1989)
147. N. Meyer-Vernet, A. Lecacheux, B.M. Pedersen, Icarus 123, 113 (1996)
148. N. Meyer-Vernet, A. Lecacheux, B.M. Pedersen, Icarus 132, 311-320 (1998)
149. N. Meyer-Vernet, A. Lecacheux, M.L. Kaiser, D.A. Gurnett, Geophys. Res. Letters 36, 3103 (2009a)
150. N. Meyer-Vernet, M. Maksimovic, A. Czechowski, I. Mann, I. Zouganelis, K. Goetz, M.L. Kaiser, J.-L. Bougeret, S.D. Bale, Dust detection by the wave instrument on Stereo: nanoparticles picked-up by the solar wind?, Solar Phys. 256, 463-474 (2009b)
151. T. Minato, M. Köhler, H. Kimura, I. Mann, T. Yamamoto, Astron. Astrophys. 424, L13-L16 (2004)
152. T. Minato, M. Köhler, H. Kimura, I. Mann, T. Yamamoto, Astron. Astrophys. 452, 701 (2006)
153. N.Y. Misconi, J. Geophys. Res. 98, 18 951-18 961 (1993)
154. J.D. Mitchell, D.J. Walter, C.L. Croskey, R.A. Goldberg, ESA SP-370, 95-100 (1995)
155. J.D. Mitchell, C.L. Croskey, R.A. Goldberg, Geophys. Res. Letters 28 no 8, 1423-1426 (2000)
156. G.E. Morfill and E. Grün, Planet. Space Sci. 27, 1269-1292 (1979)
157. T. Mukai, Astron. Astrophys. 99, 1 (1981)
158. T. Mukai and G. Schwehm, Astron. Astrophys. 95, 373-382 (1981)
159. T. Mukai and T. Yamamoto, Astron. Astrophys. 107, 97-100 (1982)
160. S. Mukai, I. Sano, M. Yasumoto, Adv. Space Res. 32, 2181-2190 (2003)
161. T.G. Northrop, Adiabatic Motion of Charged Particles, Wiley, NewYork (1963)
162. P. Oberc, Icarus 111, 211 (1994)
163. P. Oberc, Adv. Space Res 17, 110 (1996)
164. E.J. Öpik, Physics of Meteor Flight in the Atmosphere, (Interscience Publishers, New York, 1958), p. 174
165. M.M. Oppenheim, L.P. Dyrud, A.F. vom Endt, J. Geophys. Res. 108, 1064 (2003)
166. E.N. Parker, Astrophys. J. 128, 664 (1958)
167. W. Paul and H. Steinwedel, Apparatus for separating charged particles of different specific charges, Patent 944,900, German Patent, 1956.
168. J. Pavlu, A. Velyhan, I. Richterova, Z. Nemecek, J. Safrankova, I. Cermak, and P. Zilavy, IEEE Trans. Plasma Sci. 32(2), 704-708 (2004).
169. J. Pavlu, A. Velyhan, I. Richterova, J. Safrankova, Z. Nemecek, J. Wild, and M. Jerab, Vacuum 80 (6), 542547 (2006).
170. J. Pavlu, I. Richterova, Z. Nemecek, J. Safrankova, J. Wild, IEEE Trans. Plasma Sci., 35 (2), 297-302, (2007).
171. J. Pavlu, I. Richterova, Z. Nemecek, J. Safrankova, and I. Cermak, Faraday Discuss. 137, 139-155 (2008).
172. J. Pavlu, J. Safrankova, Z. Nemecek, and I. Richterova, Contrib. Plasma Phys. 49 (3), 169186 (2009).
173. A. Pellinen–Wannberg, Atmos. Chem. Phys. 4, 649-655 (2004)
174. A. Pellinen–Wannberg and G. Wannberg, J. Geophys. Res. 99, 11379-11390 (1994)
175. A. Pellinen–Wannberg ans G. Wannberg, Journal of Atmospheric and Terrestrial Physics 58, 495-506 (1996)
176. A. Pellinen–Wannberg, G. Wannberg, J.Kero, C. Szasz and A. Westman, URSI Radio Scienc Bulletin 324 (2008)
177. A. Pellinen–Wannberg, E. Murad, B. Gustavsson, U. Brändström, C.-F. Enell, C. Roth, I.P. Williams, Å. Steen, Geophys. Res. Lett. 31 (2004) doi: 10.1029/2003GL018785
178. A. Pellinen-Wannberg, E. Murad, N. Brosch, I. Haggstrom, T. Khayrov, in (J.A. Fernandez, D. Lazzaro, D. Prialnik R. Schulz, eds.), Icy Bodies of the Solar System Proceedings IAU Symposium No. 263, International Astronomical Union, doi:10.1017/S1743921310001869 (2010)
179. B. Peucker-Ehrenbrink, Geochim. Cosmochim. Acta 60, 3187-3196 (1996)
180. J.M.C. Plane, Annales Geophysicae 18, 807-814 (2000)
181. J.M.C. Plane, in Meteors in the Earth' s Atmosphere, ed. By E. Murad, E. and I.P. Williams (Cambridge University Press, 2002), pp. 289-309
182. J.M.C. Plane, C.S. Gardner, J. Yu, et al. J. Geophys. Res. 104, 3773-3788 (1999)
183. Y. V. Platov, S. A Chernouss, M. J. Kosch, J. Spacecraft Rockets 41, 667 (2004)

184. O.P. Popova, Earth, Moon, and Planets 95, 303-319 (2004)
185. O.P. Popova, A.S. Strelkov, S.N. Sidneva, Adv. Space Res. 39, 357 (2007)
186. F. Postberg, S. Kempf, J.K. Hillier, R. Srama, S.F. Green, N. McBride, E. Grün, Icarus 193, 438-454 (2008)
187. F. Postberg, S. Kempf, J. Schmidt, N. Brilliantov, A. Beinsen, B. Abel, U. Buck, and R. Srama, Nature 459 (7250), 1098-1101 (2009).
188. B.R. Ragot and S.W. Kahler, Astrophys. J. 594, 1049 - 1059 (2003)
189. N.N. Rao, P.K. Shukla, M.Y. Yu, Planet. Space Sci. 38, 543 (1990)
190. M. Rapp, Ann. Geophys. Vol. 27, 2417-2422 (2009)
191. M. Rapp and F.J. Lübken, Atmos. Chem. Physics 4, 2601 (2004)
192. M. Rapp, F.J. Lubken, T.A. Blix, Adv. Space Res. 31, 2033 (2003)
193. M. Rapp, I. Strelnikova, B. Strelnikov, et al., Joun. Geophys, Res. 115, D00I16 (2010)
194. M. Rapp, J. Hedin, I. Strelnikova, et al., Geophys. Res. Lett. 32, L23821 (2005)
195. I. Richterova, Z. Nemecek, J. Safrankova, and J. Pavlu, IEEE Trans. Plasma Sci. 32(2), 617-622 (2004).
196. I. Richterova, J. Pavlu, Z. Nemecek, and J. Safrankova, Phys. Rev. B 74(23), 235430 (2006a).
197. I. Richterova, J. Pavlu, Z. Nemecek, J. Safrankova, and P. Zilavy, Adv. Sp. Res. 38(11), 2551-2557 (2006b).
198. I. Richterova, Z. Nemecek, J. Safrankova, J. Pavlu, and M. Beranek, IEEE Trans. Plasma Sci. 35(2), 286-291 (2007).
199. I. Richterova, M. Beranek, J. Pavlu, Z. Nemecek, and J. Safrankova, Icarus, in press, 2010a.
200. I. Richterova, M. Beranek, J. Pavlu, Z. Nemecek, and J. Safrankova, Phys. Rev. B 81(7), 075406 (2010b).
201. F.J.M. Rietmeijer in: (Murad E and Williams I P) *Meteors in the Earth's atmosphere* (Cambridge: Cambridge University Press) p 215 (2002)
202. R. Roble, in Energetics of the mesosphere and thermosphere in AGU Monograph, The Upper Mesosphere and Lower Thermoshphere: A review of experiment and theory. 87, ed By R. M. Johnson, T. M. Killeen (American Geophysical Union, Washington DC, 1995), pp. 1-20
203. L.A. Rogers, K.A. Hill, R.L. Hawkes, Planet. Space Sci. 53, 1341-1354 (2005)
204. M. Rosenberg, Planet. Space Sci. 41, 229 (1993)
205. M. Rosenberg, Planet. Space Sci. 56, 1190-1193 (2008)
206. M. Rosenberg and V.W. Chow, Planet. Space Sci. 46, 103 (1998)
207. M. Rosenberg and P.K. Shukla, J. Geophys. Res. 105, 23135 (2000)
208. M. Rosenberg and P.K. Shukla, Planet. Space Sci. 50, 261 (2002)
209. M. Rosenberg and G. Sorasio, J. Spacecr. Rockets 43, 245 (2006)
210. M. Rosenberg and R.L. Merlino, Planet. Space Sci. 55, 1464 (2007)
211. M. Rosenberg, M. Salimullah, R. Bharuthram, Planet. Space Sci. 47, 1517 (1999)
212. J. Rosinski and R.H. Snow, J. Meteorol. 18, 736 (1961)
213. C. T. Russell, L. K. Jian, H. R. Lai, T.L. Zhang, A. Wennmacher, J. G. Luhmann, in: (M. Maksimovic, K. Issautier, N. Meyer-Vernet, M. Moncuquet, F. Pantelliniet Eds.) AIP CP 1216, 522-525 (2010).
214. W.A. Scales and G. Ganguli, New J. Phys. 6, 12 (2004)
215. R.W. Schunk and A.F. Nagy Ionospheres: Physics, Plasma Physics, and Chemistry (Cambridge University Press, 2000)
216. N.A. Schwadron, J. Geiss, L.A Fisk, G. Gloeckler, T.H. Zurbuchen, R. von Steiger, J. Geophys. Res. 105, 7465 (2000)
217. Schulte and Arnold (1992)
218. H. Seiler, J. Appl. Phys. 54, R1 (1983).
219. P.K. Shukla and V.P. Silin, Phys. Scripta 45, 508 (1992)
220. P.K. Shukla and A.A. Mamun, Introduction to Dusty Plasma Physics, (IOP Publ., Bristol, 2002)
221. A. Sickafoose, J. Colwell, M. Hornyi, and S. Robertson, J. Geophys. Res. 106(A5), 8343 (2001).
222. J. Spann, M. Abbas, C. Venturini, and R. Comfort, Phys. Scr. T89, 149-153 (2001).
223. P. Spurný, J. Borovička, P. Koten, Earth, Moon, and Planets 95, 569-578 (2004)
224. O.C. St.Cyr, S.P. Plunkett, D.J. Michels, S.E. Paswaters, M.J. Koomen, G.M. Simnett, B.J. Thompson, J.B. Gurman, R. Schwenn, D.F. Webb, E. Hildner, P.L. Lamy, J. Geophys. Res. 105,18169-18186 (2000)
225. E. Sternglass, Scientific Paper 6-94410-2-P9, Westinghouse Research Laboratories, Pittsburgh 35 (1957).
226. Z. Sternovsky, Z. Nemecek, J. Safrankova, and A. Velyhan, IEEE Trans. Plasma Sci. 29 (2), 292-297 (2001).
227. D. M. Suszcynsky and J. E. Borovsky, Phys. Rev. A 45, 6424, (1992).
228. D. M. Suszcynsky, J. E. Borovsky, and C. K. Goertz, J. Geophys., Res., Solid Earth 97, 2611 (1992).
229. J. Svestka, I. Cermak, and E. Grün, Adv. Sp. Res. 13(10), 199-202 (1993).
230. C. Szasz, J. Kero, A. Pellinen–Wannberg, et al., Earth, Moon, and Planets 95 (1-4), 101-107 (2004)

231. C. Szasz, J. Kero, A. Pellinen–Wannberg, D.D. Meisel, G. Wannberg, A. Westman, Earth, Moon, and Planets doi:10.1007/s11038-007-9206-y (2007)
232. C. Szasz, J. Kero, D.D. Meisel, A. Pellinen–Wannberg, G. Wannberg, A. Westman, Monthly Notices of the Royal Astronomical Society 388, 15-25 (2008)
233. S. Taylor, J. H. Lever, R. P. Harvey, Nature 392, 899-903 (1998)
234. A. G. G. M. Tielens, C.F. McKee, C.G, Seab, and D. J. Hollenbach, D. J. Astrophys. J., 431, 321-340(1994).
235. D. Tsintikidis, D. Gurnett, L.J. Granroth, et al., J. Geophys. Res. 99, 2261 (1994)
236. V.M. Vasyliunas and G.L. Siscoe, J. Geophys. Res. 81, 1247-1252 (1976)
237. A. Velyhan, P. Zilavy, J. Pavlu, J. Safrankova, and Z. Nemecek, Vakuum 76 (4), 447-455 (2004)
238. F. Verheest, Waves in dusty space plasmas, Kluwer Academic Publishers, Dordrecht (2000)
239. D. Vinkovic, Adv.Space Res. 39, 574-582 (2007)
240. T. Vondrak, J.M.C. Plane, S. Broadley, D. Janches. Atmos. Chem. Phys. 8, 7015-7031 (2008)
241. B. Walch, M. Horanyi, and S. Robertson, IEEE Trans. Plasma Sci. 22(2), 97-102 (1994)
242. B. Walch, M. Hornyi, and S. Robertson, Phys. Rev. Lett. 75(5), 838-841 (1995)
243. G. Wannberg, A. Pellinen–Wannberg, A. Westman, Radio Science 31, 497-518 (1996)
244. Y. Watanabe, J. Phys. D: Appl. Phys. 39, R329R361 (2006)
245. A. Wehry and I. Mann, Astron. Astrophys. 341, 296-303 (1999)
246. E.C. Whipple, Rep. Prog. Phys. 44, 1197 (1981)
247. D.M. Wood, Phys. Rev. Lett. 46, 749 (1981)
248. A.M. Zadorozhny, A.A. Tyutin, G. Witt, et al., Geophys. Res. Lett. 20, 2299 (1993)
249. Q.H. Zhou and M.C. Kelley, J. Atm. Solar Terr. Phys. 59, 739 (1997)
250. P. Zilavy, Z. Sternovsky, I. Cermak, Z. Nemecek, and J. Safrankova, Vacuum 50(1-2), 139-142 (1998).